



**HAL**  
open science

# Nonparametric estimation for additive concurrent regression models

Elodie Brunel, Fabienne Comte, Céline Duval

► **To cite this version:**

Elodie Brunel, Fabienne Comte, Céline Duval. Nonparametric estimation for additive concurrent regression models. 2024. hal-04600808

**HAL Id: hal-04600808**

**<https://hal.science/hal-04600808>**

Preprint submitted on 4 Jun 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# NONPARAMETRIC ESTIMATION FOR ADDITIVE CONCURRENT REGRESSION MODELS

E. BRUNEL<sup>(1)</sup>, F. COMTE<sup>(2)</sup>, C. DUVAL<sup>(3)</sup>

**ABSTRACT.** We consider an additive functional regression model where the responses are  $N$  *i.i.d.* one-dimensional processes  $(Y_i(t), i = 1, \dots, N)$  and the  $K$  explanatory random processes  $X_{i,j}(t)$  for  $j = 1, \dots, K$  are observed for  $t \in [0, \tau]$ ,  $\tau$  being fixed. The coefficients in the model are  $K$  unknown functions  $t \mapsto b_j(t)$  for  $j = 1, \dots, K$  and we build nonparametric least squares estimators under several general settings of explanatory processes, for example, continuous or inhomogeneous counting processes. We bound a mean-square type risk of the estimators from which rates of convergence are deduced. Optimality of the rates is established. An adaptive procedure is then tailored and proved to lead to relevant anisotropic model selection, simultaneously for all functions. Numerical illustrations and a real data example show the practical interest of the theoretical strategy. June 4, 2024

*Keywords and phrases:* Adaptive estimation. Continuous observation. Functional data. Least squares estimator. Nonparametric regression function estimation. Projection method.

## 1. INTRODUCTION

Everybody knows everything about the linear regression model. A natural setting extending it, is when the variable  $Y_i$  to be explained as well as the explanatory variables  $(X_{i,j})_{1 \leq j \leq K}$  for individuals  $i = 1, \dots, N$  are no longer real valued but all random functions of time. In other words, the observations become  $N$  independent paths of  $K$  random processes over some time interval  $[0, \tau]$ ,  $\tau$  being fixed. The fixed coefficients of the regression model can be modelled as deterministic functions of time, for which nonparametric estimation can be conducted in the spirit of the least-squares paradigm. Formally, we consider the functional regression model

$$(1) \quad Y_i(t) = \sum_{j=1}^K b_j(t) X_{i,j}(t) + \sigma(t, \mathbf{X}_i(t)) \varepsilon_i(t), \quad t \in [0, \tau], \quad i = 1, \dots, N,$$

For  $i = 1, \dots, N$ , the processes  $(\varepsilon_i(t))_{t \in [0, \tau]}$  are centered, independent and identically distributed in  $i$ , with  $\mathbb{E}(\varepsilon_1^2(t)) = 1$ ; the processes  $(\mathbf{X}_i(t) = (X_{i,1}(t) \dots X_{i,K}(t)))_{t \in [0, \tau]}$  are *i.i.d.* and independent of the  $(\varepsilon_i(t))_{t \in [0, \tau]}$ . The integer  $K$  is assumed to be fixed, small compared to  $N$ , and known; it is the number of explanatory variables. The homoskedastic case where  $\sigma(t, \mathbf{x}) = \sigma$  is a natural particular case. The functions  $b_1(t), \dots, b_K(t)$  are deterministic, unknown and to be estimated.

These models, known as *concurrent regression models*, are described and studied from numerical point of view, in Chapter 14 of Ramsay and Silverman (2005), [25]. They are also defined as *varying-coefficient linear models*, and they have received a lot of attention in the past two decades. Various fields of applications are targeted; survival analysis (see Teodorescu *et al.* (2010), [26], Li

---

<sup>(1)</sup>: Université de Montpellier, IMAG UMR CNRS 5149, FRANCE, email: elodie.brunel-piccinini@umontpellier.fr

<sup>(2)</sup>: Université Paris Cité, CNRS, MAP5 UMR 8145, F-75006, FRANCE, email: fabienne.comte@u-paris.fr

<sup>(3)</sup>: Université de Lille, CNRS UMR 8524, Laboratoire Paul Painlevé, Lille, FRANCE. email: celine.duval@univ-lille.fr.

*et al.* (2016), [22]), functional linear regression for longitudinal data or repeated measurements (see Yao *et al.* (2005a, b), [28]-[29], Huang *et al.* (2002), [19]), or nonparametric regression in varying-coefficient models (see Eubank *et al.* (2004), [12], Fan *et al.* (2003), [13]), more recently functional ridge regression (see Manrique *et al.* (2018), [24]), and functional regression for non euclidean responses such as fMRI signals (see Bhattacharjee and Müller (2022), [2]), hypothesis testing or variable selection in time-varying coefficient models (see Ghosal and Maity (2022a,b), [15]-[16]), or non linear time series modelling in econometrics (see Li *et al.* (2020), [21]).

The context the nearest of ours is the model studied in Manrique *et al.* (2018), [24]: they consider Model (1) in the particular setting where  $K = 1$  and  $X_{i,1}$  being a centered process. They propose a functional ridge regression estimator of the function  $b_1$  and obtain consistency and rates of convergence results under general assumptions allowing for non compact support and non continuous explanatory variable (their condition A5a seems even to ask for specifically low regularity). However, these conditions are technical, difficult to interpret and generalize, and the optimality of the procedure is rather uncertain.

Chagny *et al.* (2023), [6] consider a functional linear model formulated as an inverse problem where taking a Dirac function as kernel of their slope operator matches our model for  $K = 1$ . Their one dimensional explanatory process  $X$  is supposed to admit a Karhunen-Loeve decomposition and to satisfy technical assumptions (see their assumptions  $\mathcal{A}_2$  to  $\mathcal{A}_6$ , for instance fulfilled by gaussian processes). Their estimation strategy consists in estimating the eigenvalues and eigenfunctions of the decomposition, which is quite different from our methodology where the estimation basis is user-chosen. We have in common with their approach, the least-square contrast paradigm to build projection estimators. A novelty of our strategy is to exploit tools from statistical inference for stochastic processes, see Comte and Genon-Catalot (2024), [8], rather than those standardly used in functional data analysis.

We highlight the following contributions of this work:

- (1) We take advantage of methods usually dedicated for statistical inference for stochastic processes (specifically for SDEs) and adapt them to functional data analysis.
- (2) We consider a compact support setting which allows to handle simply general and different types of explanatory processes, which can be continuous or not.
- (3) We propose direct and simultaneous estimation of  $K$  functions, with a simple and fast estimation method, including anisotropic model selection.
- (4) We obtain rates of convergence proved to be optimal.
- (5) We emphasize that, thanks to the additive feature of the model, we do not face any curse of dimensionality. The theoretical global rate is related to the smallest regularity among those of the functions  $b_k$ ,  $k = 1, \dots, K$ . Though, numerical experiments show that the estimation algorithm chooses different orders of developments for each of the functions.

Let us describe more precisely the content of the paper. We build a nonparametric estimator of the  $K$ -dimensional function  $t \in [0, +\infty) \mapsto (b_j(t), j = 1, \dots, K) \in \mathbb{R}^K$  from the continuous time observation of the  $N$  sample paths of the  $(Y_i(t))_{t \in [0, \tau]}$  and  $(X_{i,j}(t))_{t \in [0, \tau], 1 \leq j \leq K}$ , throughout a fixed time interval  $[0, \tau]$ . The asymptotic framework is  $N \rightarrow +\infty$  and  $\tau$  fixed. For  $\mathbf{m} = (m_1, \dots, m_K) \in \mathbb{N}^K$ , we consider a space  $S_{\mathbf{m}} = S_{m_1} \times \dots \times S_{m_K}$  defined as a product of  $K$  finite-dimensional subspaces of  $\mathbb{L}^2([0, \tau])$  with respective dimensions  $m_j$ . For each  $\mathbf{m}$ , we define simultaneously projection estimators of all the functions  $b_j$ ,  $j = 1, \dots, K$ . The estimators are defined, the required assumptions are stated and explained in Section 2. The rates of the estimators and their optimality on regularity spaces is established in Section 3. The prediction risk is also studied. In Section 4, we show that the estimator can be seen as minimizing a projection contrast, in a least-squares spirit fitted to the context. This allows us to propose a model selection

device and to define a data-driven choice of  $\mathbf{m}$ , when  $\sigma(t, \mathbf{x})$  is assumed to be uniformly bounded. The resulting estimator is proved to be adaptive, in the sense that it reaches an automatic and non-asymptotic squared bias-variance compromise. Simulation experiments presented in Section 5 show that the method is simple to implement, and provides well-performing estimators, relying on fast numerical procedures. A real data example relying on a recent electricity data set is given. The proofs of the main results are given in Section 6. The proofs of the technical and auxiliary results are deferred to Supplementary Material.

**Notation.** We end this section with some of the main notations used in the article. For  $A$  a matrix, we denote by  $A^T$  the transpose of  $A$  and by  $\|A\|_{\text{op}}$  the operator norm of  $A$ , that is the square root of the largest eigenvalue of  $AA^T$ . If  $A$  is a square matrix,  $\text{Tr}(A)$  denotes the trace of  $A$  and if  $A$  is symmetric,  $\|A\|_{\text{op}} = \sup\{|\lambda_i|\}$  where  $\lambda_i$  are the eigenvalues of  $A$ . If, in addition,  $A$  is invertible,  $\|A^{-1}\|_{\text{op}} = 1/\min\{|\lambda_i|\}$ .

For  $h \in \mathbb{L}_\tau^2 = \mathbb{L}^2([0, \tau])$ , we denote by  $\|h\| = (\int_0^\tau h^2(t)dt)^{1/2}$  its  $\mathbb{L}^2$ -norm and  $\|\mathbf{x}\|_{2,r}$  denotes the Euclidian norm of the vector  $\mathbf{x} = (x_1, \dots, x_r)^T$  of  $\mathbb{R}^r$ . For  $\mathbf{h}(t) = (h_1(t), h_2(t), \dots, h_K(t))^T$  and  $\mathbf{g}(t) = (g_1(t), \dots, g_K(t))^T$  elements of  $(\mathbb{L}_\tau^2)^K = \mathbb{L}_\tau^2 \times \dots \times \mathbb{L}_\tau^2$ , we set  $\|\mathbf{h}\| = (\sum_{k=1}^K \int_0^\tau h_k^2(t)dt)^{1/2}$  and  $\langle \mathbf{h}, \mathbf{g} \rangle = \sum_{k=1}^K \int_0^\tau h_k(t)g_k(t)dt$  for respectively the  $\mathbb{L}^2$ -norm and the scalar product of  $(\mathbb{L}_\tau^2)^K$ .

Denote by  $\mathbf{X}_i(t) = (X_{i,j}(t))_{1 \leq j \leq K}$  a  $K \times 1$  vector and  $\mathbf{Y}(t) = (Y_i(t))_{1 \leq i \leq N}$  a  $N \times 1$  vector. With these notations, Model (1) can be written for  $\mathbf{b}(t) = (b_j(t))_{1 \leq j \leq K}$ ,

$$(2) \quad Y_i(t) = \mathbf{b}(t)^T \mathbf{X}_i(t) + \sigma(t, \mathbf{X}_i(t))\varepsilon_i(t), \quad t \in [0, \tau], \quad i = 1, \dots, N.$$

We also introduce two specific semi-norms, that are norms under assumption  $(\mathcal{A}_S)$  considered below. We introduce for  $N \geq 1$  and  $t \in [0, \tau]$ , the  $K \times K$  nonnegative symmetric matrices  $\Gamma_N(t), \Gamma(t)$  given by:

$$(3) \quad \Gamma_N(t) = \left( \frac{1}{N} \sum_{i=1}^N X_{i,j}(t)X_{i,k}(t) \right)_{1 \leq j, k \leq K}, \quad \Gamma(t) = (\mathbb{E}[X_{1,j}(t)X_{1,k}(t)])_{1 \leq j, k \leq K}.$$

Note that  $\Gamma(t) = \mathbb{E}[\Gamma_N(t)] = \mathbb{E}[\mathbf{X}_1(t)^T \mathbf{X}_1(t)]$ . For  $\mathbf{h} = (h_1, \dots, h_K)^T \in (\mathbb{L}_\tau^2)^K$ , we set

$$(4) \quad \|\mathbf{h}\|_N^2 = \int_0^\tau \mathbf{h}(t)^T \Gamma_N(t) \mathbf{h}(t) dt, \quad \|\mathbf{h}\|_\Gamma^2 = \int_0^\tau \mathbf{h}(t)^T \Gamma(t) \mathbf{h}(t) dt.$$

It is easy to see that  $\|\cdot\|_N$  and  $\|\cdot\|_\Gamma$  are semi-norms. We show further that they are norms if  $\Gamma_N(t)$  and  $\Gamma(t)$  are invertible for all  $t \in [0, \tau]$ . Then, the associated scalar products, for vector functions  $\mathbf{g} = (g_1, \dots, g_K)^T$  and  $\mathbf{h} = (h_1, \dots, h_K)^T$ , are

$$\langle \mathbf{g}, \mathbf{h} \rangle_N = \int_0^\tau \mathbf{g}(t)^T \Gamma_N(t) \mathbf{h}(t) dt, \quad \langle \mathbf{g}, \mathbf{h} \rangle_\Gamma = \int_0^\tau \mathbf{g}(t)^T \Gamma(t) \mathbf{h}(t) dt.$$

## 2. DEFINITION OF ESTIMATORS OF $b_j$ , FOR $j = 1, \dots, K$

**2.1. Projection estimator.** Consider  $(\varphi_j, j \geq 1)$  a basis of  $\mathbb{L}_\tau^2$  composed of measurable functions of  $\mathbb{L}_\tau^2$  such that  $\int_0^\tau \varphi_j^2(x)dx \leq 1$ . Note that this is not necessarily an orthonormal basis. Let  $S_m$  be the subspace linearly spanned by  $(\varphi_j, 1 \leq j \leq m)$ . We build estimators of  $(b_1, \dots, b_K)$  on  $S_{m_1} \times \dots \times S_{m_K} := S_{\mathbf{m}}$ , with  $\mathbf{m} = (m_1, \dots, m_K)$ . This is equivalent to estimate the coefficients  $\hat{\beta}_{j,k}$  of each  $b_k$  in the basis. Set

$$(5) \quad \hat{\mathbf{b}}_{\mathbf{m}}(t) = (\hat{b}_1(t), \dots, \hat{b}_K(t))^T \quad \text{where} \quad \hat{b}_k(t) = \sum_{j=1}^{m_k} \hat{\beta}_{k,j} \varphi_j(t),$$

such that  $\widehat{B}_{\mathbf{m}} = (\widehat{\beta}_{1,1}, \dots, \widehat{\beta}_{1,m_1}, \widehat{\beta}_{2,1}, \dots, \widehat{\beta}_{2,m_2}, \dots, \widehat{\beta}_{K,1}, \dots, \widehat{\beta}_{K,m_K})^T$  is solution of

$$\widehat{\Psi}_{\mathbf{m}} \widehat{B}_{\mathbf{m}} = \widehat{V}_{\mathbf{m}}, \quad \text{with} \quad \widehat{V}_{\mathbf{m}} = (\widehat{V}_{1,m_1}^T, \dots, \widehat{V}_{K,m_K}^T)^T,$$

and  $\widehat{V}_{j,m_j}$  are  $m_j \times 1$  vectors,  $j = 1, \dots, K$ , given by

$$\widehat{V}_{j,m_j} = \left( \frac{1}{N} \int_0^\tau \varphi_p(t) \sum_{i=1}^N X_{i,j}(t) Y_i(t) dt \right)_{1 \leq p \leq m_j}^T.$$

Moreover, for

$$(6) \quad |\mathbf{m}| := m_1 + \dots + m_K,$$

$\widehat{\Psi}_{\mathbf{m}}$  is the  $|\mathbf{m}| \times |\mathbf{m}|$  symmetric matrix with blocks  $\widehat{\Psi}_{m_j, m_k}$  of size  $m_j \times m_k$ :

$$(7) \quad \widehat{\Psi}_{\mathbf{m}} = \begin{pmatrix} \widehat{\Psi}_{m_1, m_1} & \dots & \widehat{\Psi}_{m_1, m_K} \\ \vdots & & \vdots \\ \widehat{\Psi}_{m_K, m_1} & \dots & \widehat{\Psi}_{m_K, m_K} \end{pmatrix}, \quad \widehat{\Psi}_{m_j, m_k} = \left( \int_0^\tau \varphi_p(t) \varphi_q(t) [\Gamma_N(t)]_{j,k} dt \right)_{\substack{1 \leq p \leq m_j \\ 1 \leq q \leq m_k}}.$$

We also define the a.s. limit of  $\widehat{\Psi}_{\mathbf{m}}$ ,  $\Psi_{\mathbf{m}} = \mathbb{E}[\widehat{\Psi}_{\mathbf{m}}]$ :

$$\Psi_{\mathbf{m}} = (\Psi_{m_j, m_k})_{1 \leq j, k \leq K}, \quad \Psi_{m_j, m_k} = \left( \int_0^\tau \varphi_p(t) \varphi_q(t) [\Gamma(t)]_{j,k} dt \right)_{\substack{1 \leq p \leq m_j \\ 1 \leq q \leq m_k}}.$$

And we denote by  $\Theta_{\mathbf{m}} = (\Theta_{m_j, m_k})_{1 \leq j, k \leq K}$  the  $|\mathbf{m}| \times |\mathbf{m}|$  symmetric matrix built similarly to  $\Psi_{\mathbf{m}}$ , but given by the blocks  $m_j \times m_k$

$$(8) \quad \Theta_{m_j, m_k} = \left( \int_0^\tau \int_0^\tau \varphi_p(t) \varphi_q(s) \mathbb{E}[X_{1,j}(t) X_{1,k}(s) \sigma(t, \mathbf{X}_1(t)) \sigma(s, \mathbf{X}_1(s))] \mathbb{E}[\varepsilon_1(t) \varepsilon_1(s)] dt ds \right)_{\substack{1 \leq p \leq m_j \\ 1 \leq q \leq m_k}}.$$

We explain in Section 4.2 why  $\widehat{B}_{\mathbf{m}}$  is a least-squares estimator. Then, if the matrix  $\widehat{\Psi}_{\mathbf{m}}$  is invertible, the estimator (5) can be computed from the coefficients:

$$(9) \quad \widehat{B}_{\mathbf{m}} = \widehat{\Psi}_{\mathbf{m}}^{-1} \widehat{V}_{\mathbf{m}}.$$

For the estimator to be well defined, conditions ensuring that  $\widehat{\Psi}_{\mathbf{m}}$  is invertible and that  $\Gamma_N(t)$  and  $\Gamma(t)$  are positive definite are needed, see Assumption  $(\mathcal{A}_S)$  below. Moreover for the study of the theoretical properties we consider a truncated version as follows. For constants  $\mathbf{c}_1, \mathbf{c}_2 > 0$  that can take any value, set

$$(10) \quad \Lambda_N = \{\forall t \in [0, \tau], \|\Gamma_N(t)^{-1}\|_{\text{op}} \leq \mathbf{c}_1 N^{\mathbf{c}_2}\}.$$

Using (10), we define the trimmed estimator:

$$(11) \quad \widetilde{\mathbf{b}}_{\mathbf{m}} = \widehat{\mathbf{b}}_{\mathbf{m}} \mathbf{1}_{\Lambda_N}.$$

To study the risk of the estimator with respect to  $\|\cdot\|_N$  and  $\|\cdot\|_\Gamma$ -norms, we set Assumption  $(\mathcal{A}_{X,p})$ , ensuring that these two norms are equivalent with probability near of 1. The next subsections are devoted to the presentation of these assumptions.

**2.2. Identifiability assumption.** Recall definitions (3) of  $\Gamma(t)$  and  $\Gamma_N(t)$ . Assumption  $(\mathcal{A}_S)$  below transforms all semi-norms into norms.

$(\mathcal{A}_S)$  (i)  $\forall t \in [0, \tau]$ , the matrix  $\Gamma(t)$  is invertible with

$$(12) \quad \mathfrak{f}_\Gamma := \sup_{t \in [0, \tau]} \|\Gamma(t)^{-1}\|_{\text{op}} < +\infty.$$

(ii)  $\forall t \in [0, \tau]$ ,  $\forall N \geq 1$ ,  $\Gamma_N(t)$  is a.s. invertible.

**Comments on  $(\mathcal{A}_S)$ .**

- Under  $(\mathcal{A}_S)$ ,  $\|\mathbf{h}\|_\Gamma^2 = 0$  (resp.  $\|\mathbf{h}\|_N^2 = 0$ ) implies that  $\mathbf{h}(t)^T \Gamma(t) \mathbf{h}(t) = 0$ , for almost all  $t$  in  $[0, \tau]$  (resp.  $\mathbf{h}(t)^T \Gamma_N(t) \mathbf{h}(t) = 0$ , a.s. for almost all  $t$  in  $[0, \tau]$ ). Therefore, if  $\Gamma(t)$  (resp.  $\Gamma_N(t)$ ) is invertible for all  $t \in [0, \tau]$ , we get  $\mathbf{h} \equiv 0$  in  $(\mathbb{L}_\tau^2)^K$ . It is easy to check that if  $(\mathcal{A}_S)$  (i) were not satisfied, it would be possible to express one  $X_{1,j_0}(t)$  as a linear combination of the others and it would not be possible to estimate separately the function  $b_{j_0}$ . Under  $(\mathcal{A}_S)$  (i), the  $X_{1,j}(t)$  are a.s. linearly independent.

- As  $\Gamma_N(t)$  converges a.s. to  $\Gamma(t)$  as  $N$  tends to infinity, if  $\Gamma(t)$  is invertible,  $\Gamma_N(t)$  is invertible, at least for  $N$  large enough. Hence  $(\mathcal{A}_S)$ (i) ensures  $(\mathcal{A}_S)$ (ii) for  $N$  large enough.

- Condition (12) is a uniform lower bound on the compact  $[0, \tau]$  on the smallest eigenvalue of  $\Gamma(t)$  and is not very strong under regularity assumptions on the processes. Indeed, if  $t \mapsto \Gamma(t)$  is continuous as  $\Gamma(t)$  is symmetric, positive and with real entries, Kato (1995), [20], Theorem 5.2 ensures that its  $K$  eigenfunctions are continuous implying that the smallest one is uniformly bounded away from 0 over  $[0, \tau]$ .

The link between  $(\mathcal{A}_S)$  and the invertibility of the matrix  $\widehat{\Psi}_\mathbf{m}$ , required by (9), is given by the following equality: for  $\mathbf{x} \in \mathbb{R}^{|\mathbf{m}|}$ ,

$$\mathbf{x}^T \widehat{\Psi}_\mathbf{m} \mathbf{x} = \int_0^\tau \mathbf{h}(t)^T \Gamma_N(t) \mathbf{h}(t) dt = \|\mathbf{h}\|_N^2, \quad \mathbf{x}^T \Psi_\mathbf{m} \mathbf{x} = \int_0^\tau \mathbf{h}(t)^T \Gamma(t) \mathbf{h}(t) dt = \|\mathbf{h}\|_\Gamma^2$$

where  $\mathbf{h} = (h_1, \dots, h_K)^T$  and  $h_j(t) = \sum_{k=1}^{m_j} x_{m_1 + \dots + m_{j-1} + k} \varphi_k(t)$ , and the Proposition:

**Proposition 1.** *Under  $(\mathcal{A}_S)$ , the matrices  $\widehat{\Psi}_\mathbf{m}$  and  $\Psi_\mathbf{m}$  are symmetric positive definite. Moreover, for all  $\mathbf{m}$ ,  $|\mathbf{m}| \leq N$ , it holds  $\|\Psi_\mathbf{m}^{-1}\|_{\text{op}} \leq \mathfrak{f}_\Gamma$ .*

By Proposition 1, the bound on  $\|\Gamma(t)^{-1}\|_{\text{op}}$  assumed by  $(\mathcal{A}_S)$ (i) implies a bound on  $\|\Psi_\mathbf{m}^{-1}\|_{\text{op}}$ ; the same properties hold for their empirical counterparts with large probability for  $N$  large enough, but handling this properly for any  $N$  is possible for the truncated version (11) of the estimator (5).

**2.3. Equivalence of norms and admissible explanatory processes.** Consider the following assumption:

$(\mathcal{A}_{X,p})$  (a) There exists  $p \geq 1$ , such that  $G_p^p := \max_{j=1, \dots, K} \sup_{t \in [0, \tau]} \mathbb{E}[|X_{1,j}|^p(t)] < +\infty$ .  
 (b) Suppose either  
 (i). the  $X_{1,j}(t)$ ,  $j = 1, \dots, K$ , are continuous processes such that, for all  $r > 1$  such that  $2r \leq p$ , there exists a positive constant  $\mathfrak{B}(r, \tau)$  such that  $\forall s, t \in [0, \tau]$ , with  $|t - s| \leq 1$ ,

$$(13) \quad \mathbb{E}[|X_1(t) - X_1(s)|^{2r}] \leq \mathfrak{B}(r, \tau) |t - s|^r.$$

or

(ii). For all  $j = 1, \dots, K$ , the  $X_{1,j}(t)$  have independent increments and there exist  $A_j(t)$  such that  $X_{i,j}(t) - A_j(t)$  is a centered square integrable martingale, where the function  $A_j$  is deterministic and satisfies

$$(14) \quad \max_{1 \leq j \leq K} \sup_{t \in [0, \tau]} |A_j(t)| := L_\tau < +\infty.$$

Under Assumption  $(\mathcal{A}_{X,p})(a)$ , the norm  $\|\cdot\|_\Gamma$  can be compared to the  $\mathbb{L}^2$ -norm as stated now.

**Proposition 2.** *Assume  $(\mathcal{A}_{X,p})(a)$  with  $p = 2$ , let  $\mathbf{h} \in (\mathbb{L}_\tau^2)^K$ , it holds that  $\|\mathbf{h}\|_\Gamma^2 \leq KG_2^2 \|\mathbf{h}\|^2$ . Additionally under  $(\mathcal{A}_S)$ , it holds  $\|\mathbf{h}\|^2 \leq \mathfrak{f}_\Gamma \|\mathbf{h}\|_\Gamma^2$ .*

As a consequence, Assumptions  $(\mathcal{A}_S)$  and  $(\mathcal{A}_{X,p})(a)$  for  $p = 2$  ensure that the norms  $\|\cdot\|$  and  $\|\cdot\|_\Gamma$  are equivalent for vector functions in  $\mathbb{R}^K$ .

The second assumption  $(\mathcal{A}_{X,p})(b)$  related to the explanatory processes allows two classes of continuous time processes, which may be continuous or not.

We emphasize that Assumption  $(\mathcal{A}_{X,p})(b)$  requires that all the  $X_{i,j}$  for  $1 \leq j \leq K$  are of type (i) or they are all of type (ii). Under Assumption  $(\mathcal{A}_{X,p})$ , we get  $\mathbb{E} \left[ \sup_{t \in [0, \tau]} |X_{1,j}(t)|^p \right] < +\infty$ : in case (i), it is a consequence of the Garsia Rodemich Rumsey Lemma [14] (a version of the Kolmogorov method) as stated in Jourdain and Pagès (2023), [18], see Lemma 6 in the proofs; in case (ii), it follows from the Doob maximal inequality for martingales.

### Examples of processes satisfying $(\mathcal{A}_{X,p})$ .

- Assumptions of type (i) are fulfilled by diffusion processes under regularity condition on the functional coefficients of the equation, or by classes of continuous Gaussian processes. More precisely, if  $X_{i,j}$  is solution of  $dX_{1,j}(t) = a_j(X_{1,j}(t))dt + s_j(X_{1,j}(t))dW_{1,j}(t)$ ,  $X_{1,j}(0) = \eta_{1,j}$  with  $a_j$  and  $s_j$  continuously differentiable on  $\mathbb{R}$  and  $\eta_{1,j}$  admitting moments of any order, then it fulfills  $(\mathcal{A}_{X,p})(a)$  and  $(\mathcal{A}_{X,p})(b)(i)$  for any  $p \geq 1$ , see Gloter (2000), [17].

For continuous Gaussian processes, an example is  $X_{1,j}(t) = \lambda_j \cos(a_{1,j}t)\xi_{1,j} + \mu_j \sin(a_{2,j}t)\xi_{2,j}$  for  $\xi_{1,j}$  and  $\xi_{2,j}$  independent  $\mathcal{N}(0, 1)$ . Such a process fulfills  $(\mathcal{A}_{X,p})(a)$  and  $(\mathcal{A}_{X,p})(b)(i)$  for any  $p \geq 1$ . We may also consider an explanatory process  $X_{1,j}(t)$  defined by a more general Karhunen-Loève decomposition.

- Counting processes may also be a natural idea as a choice of explanatory variables. Examples of processes fulfilling  $(\mathcal{A}_{X,p})(b)(ii)$  are given by inhomogeneous Poisson processes with intensity  $\lambda_j(\cdot)$ . Continuous diffusion processes with a deterministic drift  $a$  such as  $X_{1,j}(t) = x_{j,0} + \int_0^t a_j(u)du + \int_0^t s_j(s, X_{1,j}(s))dW_{1,j}(s)$  also fit to the condition. These two types of examples can be mixed in case  $(\mathcal{A}_{X,p})(b)(ii)$ . They fulfill  $(\mathcal{A}_{X,p})(a)$  for any  $p \geq 1$ , if the intensity  $\lambda_j$  is bounded on  $[0, \tau]$  in the inhomogeneous Poisson process case and if the functions  $a_j$  and  $s_j$  are continuous (or simply bounded) on  $[0, \tau]$  in the diffusion case.

These assumptions ensure that the empirical norm and its theoretical counterpart are equivalent for functions of  $(\mathbb{L}_\tau^2)^K$  with large probability.

**Proposition 3.** *Suppose that  $(\mathcal{A}_S)$ , and  $(\mathcal{A}_{X,p})$  for  $p \geq 2$  are satisfied, then it holds for some positive constant  $C$  that*

$$\mathbb{P} \left( \left\{ \forall \mathbf{h} \in (\mathbb{L}_\tau^2)^K, \frac{1}{2} \|\mathbf{h}\|_\Gamma^2 \leq \|\mathbf{h}\|_N^2 \leq \frac{3}{2} \|\mathbf{h}\|_\Gamma^2 \right\}^c \right) \leq CN^{-p/2}.$$

**2.4. Additional assumptions.** We set the following natural condition on the functions  $b_k$ ,  $k = 1, \dots, K$ :

( $\mathcal{A}_b$ ) The functions  $b_k : \mathbb{R}^+ \rightarrow \mathbb{R}$ ,  $k = 1, \dots, K$ , are measurable, and bounded on  $[0, \tau]$  by a constant  $\mathbf{c}_b$  (and thus belong to  $\mathbb{L}_\tau^2$ ).

Finally, an assumption on the basis is required.

( $\mathcal{A}_\varphi$ )  $\exists \omega > 0$ ,  $\exists c_\varphi > 0$  such that  $L(S_m) := \sup_{t \in [0, \tau]} \sum_{j=1}^m \varphi_j^2(t) \leq c_\varphi m^\omega$ ,  $\forall m \geq 1$ .

If the basis is orthonormal, the quantity  $L(S_m)$  was proved in Lemma 1 in Birgé and Massart (1998), [4], to be equal to  $L(S_m) = \sup_{h_1 \in S_m, \|h_1\|=1} \sup_{t \in [0, \tau]} h_1^2(t)$  where  $\|h_1\|^2 = \int_0^\tau h_1^2(t) dt$ .

Assumption ( $\mathcal{A}_\varphi$ ) holds for several classical bases of  $\mathbb{L}_\tau^2$ , with  $\omega = 1$  for the trigonometric basis on  $[0, \tau]$ ,  $\omega = 2$  for the Legendre basis on  $[0, \tau]$  and  $\omega = 1/2$  for the Hermite basis or  $\omega = 1$  for the Laguerre basis (see e.g. Comte and Marie (2023) [10]). The first two bases are orthonormal on  $[0, \tau]$  but the last one is not. We give more precise examples in Section 5.

### 3. STUDY OF ESTIMATORS OF $b_j$ , FOR $j = 1, \dots, K$ .

**3.1. Risk bound on  $\tilde{\mathbf{b}}_{\mathbf{m}}$ .** In what follows, we define the risk of our estimator as the expectation of the empirical squared norm  $\|\tilde{\mathbf{b}}_N - \mathbf{b}\|_N^2$  or the weighted squared norm  $\|\tilde{\mathbf{b}}_N - \mathbf{b}\|_\Gamma^2$ . These definitions of the risk are classically used for regression, see e.g. Baraud (2002), [1].

**Theorem 1.** *Assume that ( $\mathcal{A}_S$ ), ( $\mathcal{A}_{X,p}$ ), ( $\mathcal{A}_b$ ) and ( $\mathcal{A}_\varphi$ ) hold and that  $\mathbf{m}$  satisfies  $|\mathbf{m}| \leq N$ , where  $|\mathbf{m}|$  is given in (6). Assume moreover that*

$$(15) \quad \sup_{t \in [0, \tau]} \mathbb{E}[\varepsilon^4(t)] := \mathbf{c}_\varepsilon^2 < +\infty \text{ and } \mathfrak{M}_\sigma^2 := \sup_{j=1, \dots, K} \sup_{t \in [0, \tau]} \mathbb{E}[X_{1,j}^4(t) \sigma^4(t, \mathbf{X}_1(t))] < +\infty.$$

The estimator  $\tilde{\mathbf{b}}_{\mathbf{m}}$  of  $\mathbf{b}(t) = (b_1(t), \dots, b_K(t))^T$  satisfies, for  $c$  and  $c'$  generic constants:

• for  $p \geq (2\mathbf{c}_2 + 1 + \omega \vee 1) \vee 4$ , with  $\mathbf{c}_2$  defined in (10),

$$(16) \quad \mathbb{E}[\|\tilde{\mathbf{b}}_{\mathbf{m}} - \mathbf{b}\|_N^2] \leq \inf_{\mathbf{h} \in S_{\mathbf{m}}} \|\mathbf{h} - \mathbf{b}\|_\Gamma^2 + 2 \frac{\text{Tr}(\Psi_{\mathbf{m}}^{-1} \Theta_{\mathbf{m}})}{N} + \frac{c}{N},$$

• for  $p \geq (4\mathbf{c}_2 + 4 + \omega) \vee 8$ ,

$$(17) \quad \mathbb{E}[\|\tilde{\mathbf{b}}_{\mathbf{m}} - \mathbf{b}\|_\Gamma^2] \leq 5 \inf_{\mathbf{h} \in S_{\mathbf{m}}} \|\mathbf{h} - \mathbf{b}\|_\Gamma^2 + 4 \frac{\text{Tr}(\Psi_{\mathbf{m}}^{-1} \Theta_{\mathbf{m}})}{N} + \frac{c'}{N}.$$

Moreover, it holds that

$$(18) \quad \text{Tr}(\Psi_{\mathbf{m}}^{-1} \Theta_{\mathbf{m}}) \leq \tau \mathfrak{f}_\Gamma \mathbf{c}_\varepsilon \mathfrak{M}_\sigma \left\{ |\mathbf{m}| \wedge \left( c_\varphi \tau \sum_{j=1}^K m_j^\omega \right) \right\}.$$

We obtain a standard squared-bias variance decomposition, where (18) implies that the order of the variance is  $|\mathbf{m}|/N$  if  $\omega \geq 1$ , if  $\omega < 1$ , this order can be improved.

To conclude this section, we connect the popular risk of prediction and the integrated  $\Gamma$ -risk which is controlled in Theorem 1. Consider  $\mathbf{X}_{N+1}$  a new observation, independent from  $(Y_j, \mathbf{X}_j)_{1 \leq j \leq N}$ , for which  $Y_{N+1}$ , unobserved, is predicted by  $\hat{Y}_{N+1} = \tilde{\mathbf{b}}_{\mathbf{m}}^T \mathbf{X}_{N+1}$ .

**Proposition 4.** *Under the Assumptions of Theorem 1, it holds that*

$$\mathbb{E} \left[ \int_0^\tau \left( Y_{N+1}(t) - \tilde{\mathbf{b}}_{\mathbf{m}}^T(t) \mathbf{X}_{N+1}(t) \right)^2 dt \right] = \mathbb{E} \left[ \|\tilde{\mathbf{b}}_{\mathbf{m}} - \mathbf{b}\|_\Gamma^2 \right] + \mathbb{E} \left[ \int_0^\tau \sigma^2(t, \mathbf{X}_{N+1}(t)) dt \right].$$

Therefore, the prediction error can be controlled by the estimation error on  $\mathbf{b}$  up to a constant term: the variance of the noise term. Proposition 4 explains why the prediction error may remain large if  $\sigma$  takes large values, see the real data example in Section 5.4.



**3.2. Upper and lower bounds on rates.** To evaluate rates of convergence, we must assess the  $\mathbb{L}^2$ -norm of the estimator's bias within some regularity subspaces of  $\mathbb{L}_\tau^2$ . Such assessments are standard in nonparametric statistics, for functions  $b_j$  belonging to Sobolev spaces associated with the chosen basis.

**Proposition 5.** *Assume that for all  $j \in \{1, \dots, K\}$ , the function  $b_j$  belongs to a regularity space such that  $\inf_{h \in S_m} \|b_j - h\|^2 \leq R_j m^{-2s_j}$ . Choose  $m_j = O(N^{1/(2s_j+1)})$  and set  $s_{(1)} = \min_{j=1, \dots, K} s_j$ . Then, under the assumptions of Theorem 1 with  $\omega = 1$ , for some positive constant  $C$  it holds*

$$\mathbb{E}[\|\tilde{\mathbf{b}}_m - \mathbf{b}\|_{\Gamma}^2] \leq C N^{-2s_{(1)}/(2s_{(1)}+1)}.$$

Thus, our method has the advantage of estimating all functions simultaneously and of reaching the rate corresponding to the estimation of one function with regularity  $s_{(1)}$ , without any curse of dimensionality. The rate corresponds to the smallest regularity.

Establishing the optimality of the rate in Proposition 5 requires additional assumptions on the model. First, we need to specify the regularity classes in which optimality is studied. Therefore, we now assume that  $(\varphi_j)_{j \geq 0}$  is an orthonormal basis of  $\mathbb{L}_\tau^2$ , and define the regularity space

$$W(s, L) = \{f = \sum_{j \geq 0} \theta_j \varphi_j, \text{ such that } \forall J \geq 1, \sum_{j \geq J} \theta_j^2 \leq L^2 J^{-2s}\}.$$

Note that if  $b_j \in W(s_j, L_j)$ ,  $j \in \{1, \dots, K\}$ , it holds that  $\|b_j - b_m\|^2 \leq L_j^2 m^{-2s_j}$  and the assumption on the function  $b_j$  of Proposition 5 is fulfilled. The fact that  $b$  belongs to  $W(s, L)$  is related to its regularity. For instance, for the trigonometric basis and  $s$  integer, see Tsybakov (2009) [27] (see Definition 1.12 therein, and the sequel); this means that  $b$  is  $s$  times differentiable. This holds also for the Legendre basis, see Efromovich (1999) [11], Section 2.6. These two bases are orthonormal on a compact support and satisfy  $(\mathcal{A}_\varphi)$  with  $\omega = 1$  and  $\omega = 2$  respectively.

Second, to establish the lower bound result, we rely on the method of Tsybakov (2009) [27] based on the construction of many hypotheses. To that end, one needs to compute the distribution of  $(X(t), Y^b(t) = b(t)X(t) + \sigma(t)\varepsilon(t))_{t \in [0, \tau]}$ . Since the rate of convergence is determined by the smallest regularity, we study a lower bound result for  $K = 1$  and then derive the result for  $K \geq 1$ . A case where it is possible to specify this distribution is when the process  $\varepsilon$  is a centered Gaussian process; then, the Cameron Martin theorem (see Chagny *et al.* (2022) [6] Theorem 4, see also Lifshits (2012) [23]) can be applied. We prove the following lower bound.

**Theorem 2.** *Let  $(\varphi_j)_{j \geq 0}$  be an orthonormal basis of  $\mathbb{L}_\tau^2$ ,  $s$  and  $L$  be positive constants. Assume that  $(\mathcal{A}_S)$  and  $(\mathcal{A}_{X,p})(a)$  for  $p = 2$  hold. Let  $\varepsilon$  be a centered Gaussian process such that  $\mathbb{E}[\varepsilon^2(t)] = 1$ ,  $\forall t \in [0, \tau]$  and  $\sigma$  a bounded positive function. Denote by  $\Sigma_{\sigma\varepsilon}$  the covariance operator of the process  $(\sigma(t)\varepsilon(t))_{t \in [0, \tau]}$ , satisfying  $\|\Sigma_{\sigma\varepsilon}^{-1}\|_{\text{op}} \leq c_{\sigma\varepsilon} < \infty$  for some constant  $c_{\sigma\varepsilon}$  and for all  $b_k \in W(s_k, L_k)$  that  $\langle b_k X_{1k}, \Sigma_{\sigma\varepsilon}^{-1} b_k X_{1k} \rangle < \infty$ ,  $1 \leq k \leq K$ . Then, it holds for some positive constant  $c$ , that*

$$(19) \quad \inf_{\widehat{\mathbf{b}} = (\widehat{b}_1, \dots, \widehat{b}_K)} \sup_{b_k \in W(s_k, L_k), k=1, \dots, K} \mathbb{E}[\|\widehat{\mathbf{b}} - \mathbf{b}\|_{\Gamma}^2] \geq c N^{-\frac{2s_{(1)}}{2s_{(1)}+1}}, \quad s_{(1)} = \min_{1 \leq k \leq K} s_k.$$

The Cameron Martin theorem imposes a restriction on the set of admissible shifts  $bX$ , this is why the constraint  $\langle bX, \Sigma_{\sigma\varepsilon}^{-1} bX \rangle < \infty$  appears. In the following Corollary we provide examples where these conditions are met provided that  $bX$  is of class  $C^2$  for all  $b \in W(s, L)$ .

**Corollary 1.** *Let  $\tau = 1$ ,  $s \geq 2$  and  $L$  be positive constants. Assume  $(\mathcal{A}_S)$  and  $(\mathcal{A}_{X,p})(a)$  for  $p = 2$ . Define either  $(\sigma_1(t), \varepsilon_1(t)) = (\sqrt{t+1}, \sqrt{t+1}^{-1} W(t+1))$  or  $(\sigma_2(t), \varepsilon_2(t)) = (1, e^{-t/2} W(e^t))$ , for  $t \in [0, 1]$  and  $W$  a standard Brownian motion. Suppose, for  $k \in \{1, \dots, K\}$ ,  $X_{1k}$  is  $C^2([0, 1])$ .*

Then, it holds for some positive constant  $c$  that  $\inf_{\widehat{\mathbf{b}}} \sup_{b_k \in W(s_k, L_k)} \mathbb{E}[\|\widehat{\mathbf{b}} - \mathbf{b}\|_{\Gamma}^2] \geq cN^{-\frac{2s(1)}{2s(1)+1}}$  with  $s(1) = \min_{1 \leq k \leq K} s_k$ .

As a consequence, the rates obtained when using the trigonometric or Legendre basis on  $[0, 1]$  are minimax optimal.

#### 4. MODEL SELECTION

The choices proposed above for  $\mathbf{m}$  are asymptotic and depend on unknown regularity parameters. So, they cannot be implemented. This is why we propose a data driven model selection device. This defines a new estimator, for which we prove a non asymptotic risk bound. A preliminary step, is to show that the estimator defined in (9) is a minimum contrast estimator.

**4.1. Estimation contrast.** For  $i = 1, \dots, N$ ,  $(Y_i(t), \mathbf{X}_i(t))_{0 \leq t \leq \tau}$  from Model (2) and  $\mathbf{h}$  element of  $(\mathbb{L}_{\tau}^2)^K$ , consider the contrast

$$\begin{aligned} U_N(\mathbf{h}) &= \frac{1}{N} \int_0^{\tau} \sum_{i=1}^N [\mathbf{h}(t)^T \mathbf{X}_i(t)]^2 dt - \frac{2}{N} \sum_{i=1}^N \int_0^{\tau} Y_i(t) [\mathbf{h}(t)^T \mathbf{X}_i(t)] dt \\ (20) \quad &= \|\mathbf{h}\|_N^2 - \frac{2}{N} \int_0^{\tau} \mathbf{h}(t)^T \mathbb{X}(t)^T \mathbf{Y}(t) dt, = \|\mathbf{h}\|_N^2 - 2\langle \mathbf{h}, \mathbf{b} \rangle_N - 2\nu_N(\mathbf{h}) \end{aligned}$$

where  $\mathbb{X}(t) = (X_{i,j}(t))_{1 \leq i \leq N, 1 \leq j \leq K}$  is a  $N \times K$  matrix and  $\nu_N(\mathbf{h})$  is a centered empirical process defined by

$$(21) \quad \nu_N(\mathbf{h}) = \frac{1}{N} \sum_{i=1}^N \int_0^{\tau} [\mathbf{h}(t)^T \mathbf{X}_i(t)] \sigma(t, \mathbf{X}_i(t)) \varepsilon_i(t) dt.$$

This requires that the processes  $Y_i(t)$  and  $\mathbf{X}_i(t)$  are a.s. measurable as functions of  $t$  so that the integrals are well-defined.

The projection estimator of  $\mathbf{b}(t) = (b_1(t), \dots, b_K(t))^T$  on  $S_{\mathbf{m}}$  is given by

$$(22) \quad \widehat{\mathbf{b}}_{\mathbf{m}}(t) = (\widehat{b}_1(t), \dots, \widehat{b}_K(t))^T, \quad \widehat{\mathbf{b}}_{\mathbf{m}} = \arg \min_{\mathbf{h} \in S_{\mathbf{m}}} U_N(\mathbf{h}).$$

Standard computations setting the gradient of the least squares contrast to zero, allow to check that  $\widehat{B}_{\mathbf{m}} = (\widehat{\beta}_{1,1}, \dots, \widehat{\beta}_{1,m_1}, \widehat{\beta}_{2,1}, \dots, \widehat{\beta}_{2,m_2}, \dots, \widehat{\beta}_{K,1}, \dots, \widehat{\beta}_{K,m_K})^T$  defined in (9) minimizes (20).

To understand why  $U_N(\mathbf{h})$  provides an estimator of  $\mathbf{b}(t)$ , let us compute its expectation:

$$\begin{aligned} \mathbb{E}[U_N(\mathbf{h})] &= \mathbb{E}[\|\mathbf{h}\|_N^2] - \frac{2}{N} \mathbb{E} \left[ \int_0^{\tau} \sum_{i=1}^N \sum_{k=1}^K h_k(t) X_{i,k}(t) \sum_{k'=1}^K b_{k'}(t) X_{i,k'}(t) dt \right] \\ &= \|\mathbf{h}\|_{\Gamma}^2 - 2\mathbb{E} \left[ \int_0^{\tau} \sum_{k=1}^K h_k(t) X_{1,k}(t) \sum_{k'=1}^K b_{k'}(t) X_{1,k'}(t) dt \right] \\ &= \|\mathbf{h}\|_{\Gamma}^2 - 2\langle \mathbf{h}, \mathbf{b} \rangle_{\Gamma} = \|\mathbf{h} - \mathbf{b}\|_{\Gamma}^2 - \|\mathbf{b}\|_{\Gamma}^2. \end{aligned}$$

Obviously, since under  $(\mathcal{A}_{\mathcal{S}})$ ,  $\|\cdot\|_{\Gamma}$  is a norm,  $\mathbb{E}[U_N(\mathbf{h})]$  is minimal if  $h_j = b_j$  for  $j = 1, \dots, K$ .

**4.2. Model selection procedure.** As usual for least squares estimators, it holds that  $U_N(\widehat{\mathbf{b}}_{\mathbf{m}}) = -\|\widehat{\mathbf{b}}_{\mathbf{m}}\|_N^2$ . This quantity provides an estimate of the bias up to a term not depending on  $\mathbf{m}$ , as  $\|\mathbf{b} - \widehat{\mathbf{b}}_{\mathbf{m}}\|_N^2 = \|\mathbf{b}\|_N^2 - \|\widehat{\mathbf{b}}_{\mathbf{m}}\|_N^2$ .

For the variance term, a simpler value of the constant in (18) can be given under an additional assumption.

$(\mathcal{A}_\sigma)$  Suppose that  $\sup_{t \in [0, \tau], x \in \mathbb{R}^K} \sigma^2(t, x) := \|\sigma\|_\infty^2 < +\infty$ .

**Corollary 2.** *If  $\sigma$  satisfies  $(\mathcal{A}_\sigma)$ , then  $\text{Tr}(\Psi_{\mathbf{m}}^{-1} \Theta_{\mathbf{m}}) \leq \tau \|\sigma\|_\infty^2 |\mathbf{m}|$ .*

We take this bound as an estimate of the variance, assuming that  $\omega = 1$  in Assumption  $(\mathcal{A}_\varphi)$ . This explains why we select

$$(23) \quad \widehat{\mathbf{m}} \in \arg \min_{\mathbf{m} \in \mathcal{M}_N} \left[ U_N(\widehat{\mathbf{b}}_{\mathbf{m}}) + \text{pen}(\mathbf{m}) \right], \quad \text{pen}(\mathbf{m}) = \kappa \|\sigma\|_\infty^2 \frac{|\mathbf{m}|}{N},$$

where we consider the collection of models defined by

$$(24) \quad \mathcal{M}_N = \{ \mathbf{m} \in \{1, \dots, N\}^K, \quad \forall i \in \{1, \dots, K\}, m_i \leq N \}.$$

Note that the collection imposes  $|\mathbf{m}| \leq KN$  instead of  $N$  previously, in order to have a nesting space, that is a space containing any space of the collection:

$$\forall i \in \{1, \dots, K\}, \forall m_i \leq N, \quad S_{m_1} \times \dots \times S_{m_K} \subset S_N \times \dots \times S_N := \mathbf{S}_N.$$

Finally, we consider the estimator

$$\tilde{\mathbf{b}} = \widehat{\mathbf{b}}_{\widehat{\mathbf{m}}} \mathbf{1}_{\Lambda_N},$$

where  $\Lambda_N$  is defined by (10).

**Theorem 3.** *Assume that  $(\mathcal{A}_S)$ ,  $(\mathcal{A}_{X,p})$ ,  $(\mathcal{A}_b)$ ,  $(\mathcal{A}_\varphi)$  with  $\omega = 1$  and  $(\mathcal{A}_\sigma)$  hold, and that*

$$\sup_{t \in [0, \tau]} \mathbb{E}[e^6(t)] < +\infty.$$

*Consider the estimator  $\tilde{\mathbf{b}}$  of  $\mathbf{b}$  with any  $\widehat{\mathbf{m}}$  defined by (23). Then, there exists a numerical constant  $\kappa_0$  such for all  $\kappa \geq \kappa_0$ , it holds, for  $p \geq (2c_2 + 2) \vee 6$ ,*

$$(25) \quad \mathbb{E} \left[ \|\tilde{\mathbf{b}} - \mathbf{b}\|_N^2 \right] \leq 4 \inf_{\mathbf{m} \in \mathcal{M}_N} \left( \inf_{\mathbf{h} = (h_1, \dots, h_K)^T \in S_{\mathbf{m}}} \|\mathbf{h} - \mathbf{b}\|_\Gamma^2 + \kappa \|\sigma\|_\infty^2 \frac{|\mathbf{m}|}{N} \right) + \frac{C}{N},$$

*and for  $p \geq (4c_2 + 5) \vee 8$ ,*

$$(26) \quad \mathbb{E} \left[ \|\tilde{\mathbf{b}} - \mathbf{b}\|_\Gamma^2 \right] \leq C_1 \inf_{\mathbf{m} \in \mathcal{M}_N} \left( \inf_{\mathbf{h} = (h_1, \dots, h_K)^T \in S_{\mathbf{m}}} \|\mathbf{h} - \mathbf{b}\|_\Gamma^2 + \kappa \|\sigma\|_\infty^2 \frac{|\mathbf{m}|}{N} \right) + \frac{C'}{N},$$

*where  $C_1$  is a numerical constant,  $C$  and  $C'$  are constants depending on  $K, G, \|\sigma\|_\infty$ .*

The term  $\|\sigma\|_\infty$  is unknown, it is replaced by an estimator in the numerical experiments: the supremum over  $t$  of the least squares residuals,  $N^{-1} \sum_{i=1}^N [Y_i(t) - \tilde{\mathbf{b}}_{\mathbf{m}_N}(t)^T \mathbf{X}_i(t)]^2$  where  $\tilde{\mathbf{b}}_{\mathbf{m}_N}$  is an estimator of  $\mathbf{b}$  computed on a fixed and large enough space of the collection.

Inequality (25) means that a squared bias/variance compromise is automatically obtained for the final estimator, the bound is non asymptotic. Asymptotically, when  $\mathbf{b}$  belongs to a regularity space as described in Section 3.2, the rate given in Proposition 5 is automatically reached by the estimator.

It also follows from Theorem 3 that, for any function  $\mathbf{b}$ , there exists a numerical (universal) constant  $\kappa_0$  such that the inequality holds for all  $\kappa \geq \kappa_0$ . The numerical value of  $\kappa_0 = 24$  is found in the proofs, but it is too large in practice. For numerical implementation of the adaptive estimator, it is standard to start by preliminary simulations to obtain a relevant value for  $\kappa$ . Once the value of  $\kappa$  is chosen, it is fixed once and for all. See Baudry *et al.* (2012) [3] for the general methods of practical calibration.

5. SIMULATION AND REAL DATA

5.1. **Bases.** We consider three bases.

[T] We denote by T the trigonometric basis called "half-trigonometric" system, namely the cosine basis defined by

$$\varphi_{0,\tau}(x) = \sqrt{1/\tau} \mathbf{1}_{[0,\tau]}(t), \quad \varphi_{j,\tau}(t) = \sqrt{2/\tau} \cos(\pi j t/\tau) \mathbf{1}_{[0,\tau]}(t), \quad j = 1, \dots, m - 1,$$

see Efromovich (1999, p.46), [11]. It is clearly an orthonormal basis such that  $L(S_m^{[T]}) \leq 2/\tau m$ .

[H] We denote by H the Hermite basis, composed of Hermite functions  $h_j$  given by, for  $j \geq 0$ :

$$(27) \quad h_j(x) = c_j H_j(x) e^{-x^2/2}, \quad c_j = (2^j j! \sqrt{\pi})^{-1/2}, \quad H_j(x) = (-1)^j e^{x^2} \frac{d^j}{dx^j} (e^{-x^2}).$$

The sequence  $(h_j, j \geq 0)$  is an orthonormal basis of  $\mathbb{L}^2(\mathbb{R})$  and satisfies  $L(S_m^{[H]}) \leq C\sqrt{m}$ , see Lemma 1 in Comte and Lacour (2023), [9].

[L] We denote by L the Laguerre basis (see Comte and Genon-Catalot (2018), [7]) defined by

$$(28) \quad \ell_j(t) = \sqrt{2} L_j(2t) e^{-t} \mathbf{1}_{t \geq 0}, \quad j \geq 0, \quad L_j(t) = \sum_{k=0}^j (-1)^k \binom{j}{k} \frac{t^k}{k!}.$$

The sequence  $(\ell_j, j \geq 0)$  is an orthonormal basis of  $\mathbb{L}^2(\mathbb{R}_+)$  and satisfies  $L(S_m^{[L]}) \leq 2m$ .

The basis T is well-suited to our compactly-supported setting and is the most natural choice. We also consider the two bases L and H, which are not orthonormal on  $[0, \tau]$ , but which often work well while being parsimonious: few coefficients lead to a good estimation.

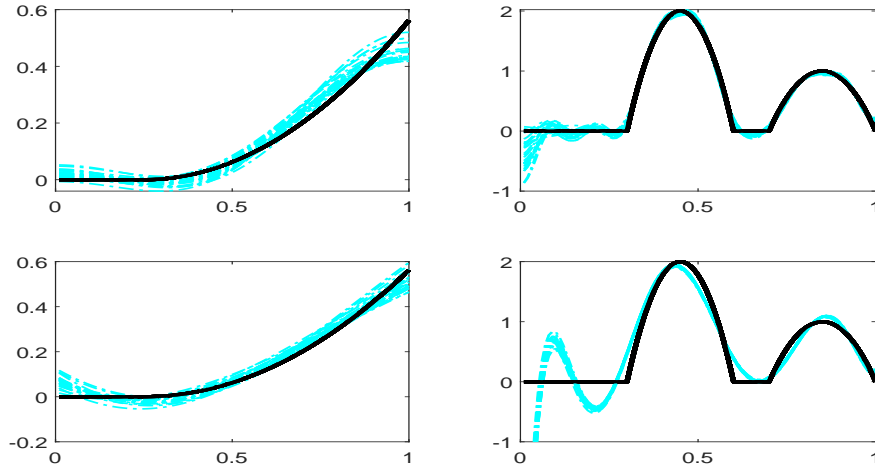


FIGURE 1. Estimation of  $b_1$  left and  $b_2$  right, with basis T (first line) and basis H (second line) for 25 paths. WASE 0.0031 for T and 0.0671 for H basis.

5.2. **An example with  $K = 2$ .** To start with, we reproduce the example of "setting 1" proposed by Manrique *et al.* (2018), [24]. They consider  $K = 2$ ,  $X_1(t) = 1$  and  $X_2(t) = \mu_X(t) + \sum_{j=1}^{10} \rho_j \xi_{i,j} \phi_j(t)$  where  $\mu_X(t) = t + \sin(t)$ , for  $j \geq 1$ ,  $\phi_j(t) = \sqrt{2} \sin((j - 1/2)\pi t)$ ,  $\rho_j =$

$1/((j-1/2)\pi)$ , and the  $\xi_{i,j}$  are i.i.d. and  $\mathcal{N}(0, 1)$ . The functions are  $b_1(t) = (t - 0.25)^2 \mathbf{1}_{[0.25, 1]}(t)$  and

$$b_2(t) = \left( -\frac{2}{0.15^2}(t - 0.45)^2 + 2 \right) \mathbf{1}_{[0.3, 0.6]}(t) + \left( -\frac{1}{0.15^2}(t - 0.85)^2 + 2 \right) \mathbf{1}_{[0.7, 1]}(t).$$

The noise is  $\sigma(t)\varepsilon(t) = c_\varepsilon \sum_{j=1}^{20} \rho_j \xi_{i,j} \phi_j(t)$  where  $c_\varepsilon \approx 7$  is chosen such that we have a similar signal to noise ratio of 2 as defined in Manrique *et al.* (2018), [24]. We compute the following weighted average squared error

$$WASE := \frac{1}{2\tau} \left[ \frac{\int_0^\tau (\widehat{b}_1(t) - b_1(t))^2 dt}{\text{range}^2(b_1)} + \frac{\int_0^\tau (\widehat{b}_2(t) - b_2(t))^2 dt}{\text{range}^2(b_2)} \right]$$

over 500 repetitions for  $N = 70$  and discrete observations  $t_j$  at 100 equispaced points in  $[0, 1]$ . The results are displayed in Table 1. The trigonometric basis T with our method performs much better than Manrique *et al.*'s proposal, but the Hermite H basis fails probably because of the lack of regularity of the functions under estimation. The mean of selected dimensions is 3 for  $b_1$  in both bases, and for  $b_2$ , 15 with the trigonometric basis and 8 for the Hermite basis. The true value of the infinite norm of  $\sigma(t)$  is used in the penalty, and the constant  $\kappa$  is set to 0.2 for both bases.

	Manrique <i>et al.</i>	Basis T	Basis H
WASE	0.0150	0.0032	0.0642
std	0.0097	0.0015	0.0088

TABLE 1. Comparison of two bases to Manrique *et al.* (2018)'s [24] results, 500 repetitions.

Figure 1 is an illustration of the quality of the results and their stability: 25 estimators of each functions are given for both bases together with the true curves. Clearly, the main problem for both bases is to recover the zero part at the beginning of the interval for the function  $b_2$ , otherwise, their performances are quite convincing. The estimator proposed by Manrique *et al.* (2018), [24] does not have this problem, but clearly fails to estimate  $b_1$  (see their Fig.1 p.997). The number of observed paths  $N = 70$  for each experiment is not very large, compared with the asymptotics.

**5.3. Simulation experiments for  $K = 3$ .** We conducted simulation experiments in the case  $K = 3$  with  $X_1(t) = 1$ ,  $X_2(t)$  a Poisson process with parameter  $\lambda = 0.5$  and  $X_3(t)$  and  $\sigma(t)\varepsilon(t)$  as the  $X_2(t)$  of Manrique *et al.* (2018), [24], described above, with here  $c_\varepsilon = 7$ . We considered the functions

$$(29) \quad b_1(t) = \cos(2.8\pi t), \quad b_2(t) = 0.25 \exp(-t/3) - 2 \exp(-2t), \quad b_3(t) = 2t^2.$$

The idea was to choose an "easy to estimate" function for basis T ( $b_1$ ) and an easy function as well for basis L ( $b_2$ ), associated with explanatory variables of different types (constant, counting process, regular Gaussian process).

The term  $\|\sigma\|_\infty$  is replaced by an estimator defined as residual least squares associated with an estimate of the function  $b_i$  for  $i = 1, 2, 3$ ,  $\widehat{b}_{m_i}$  where  $m_i = \min([\sqrt{N}/3], 5)$ . The  $X_{i,j}(t)$  are sampled at times  $t_k = k\tau/n$ , with  $\tau = 1$  and  $n = 100$ ,  $k = 1, \dots, 100$ . The penalty constants are calibrated to  $\kappa_T = 0.25$  for basis T and to  $\kappa_L = 0.125$  for basis L. The maximal dimensions considered in each collection are 12 for T and 10 for L. Larger values of these maximal dimensions have been tested, but they are never selected by the algorithm, and such restrictions make the computations much faster.

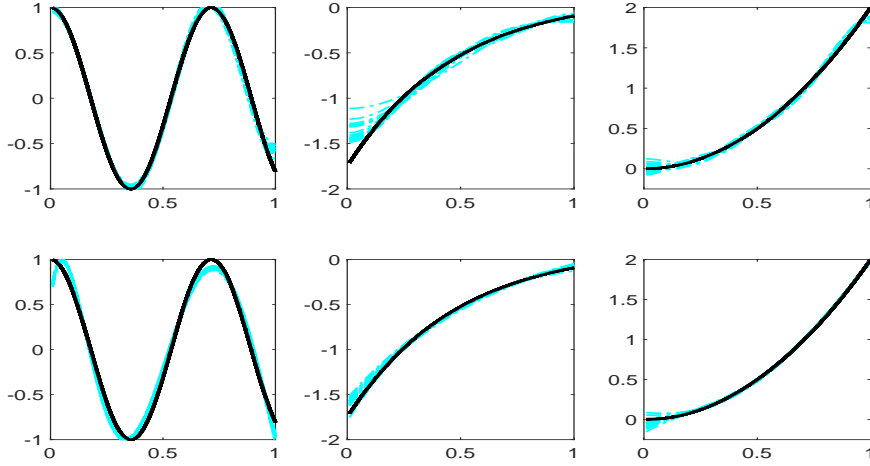


FIGURE 2. Case 1. Estimation of  $b_1$  left,  $b_2$  middle and  $b_3$  right, with basis T (first line) and basis L (second line) for 25 paths.  $100 \times \text{MISE}(b_1, b_2, b_3)$ ,  $(0.284, 0.006, 0.134)$  for T and  $(0.748, 0.001, 0.032)$  for L. Mean of the selected dimensions  $(5.6, 3.6, 5.5)$  for T and  $(6, 0, 2.0, 4.0)$  for L.

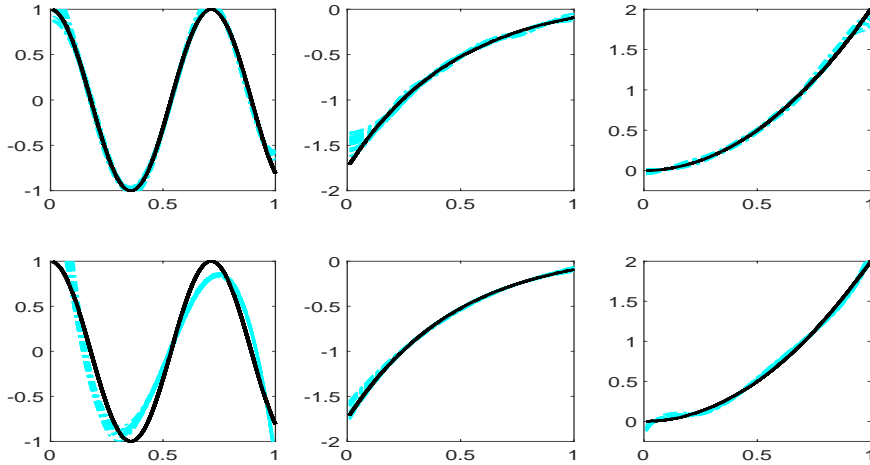


FIGURE 3. Case 2 (permuted). Estimation of  $b_1$  left,  $b_2$  middle and  $b_3$  right, with basis T (first line) and basis L (second line) for 25 paths.  $100 \times \text{MISE}(b_1, b_2, b_3)$ ,  $(0.295, 0.003, 0.137)$  for T and  $(11.00, 0.001, 0.124)$  for L. Mean of the selected dimensions  $(5.9, 4.5, 5.6)$  for T and  $(5, 0, 2.4, 6.0)$  for L.

The MISE for the estimators is here computed as follows:

$$(30) \quad \text{MISE}(b_\ell) = \frac{1}{J} \sum_{j=1}^J \frac{\tau}{n} \sum_{j=1}^n [\widehat{b}_{\ell, \widehat{m}_\ell}^{(j)}(j\tau/n) - b_\ell(j\tau/n)]^2$$

where  $\widehat{b}_{\ell, \widehat{m}_\ell}^{(j)}$  is the adaptive estimator of  $b_\ell$  for the  $j$ th simulated path,  $\ell = 1, 2, 3$ .

Function		$N = 50$		$N = 200$		$N = 1000$	
		Case 1	Case 2	Case 1	Case 2	Case 1	Case 2
$b_1$	MISE	1.678	2.063	0.274	0.323	0.074	0.092
	std	0.11	5.15	0.10	0.18	0.02	0.03
	dim	4	4.4	5.7	6.0	7.9	8.2
$b_2$	MISE	0.029	0.015	0.007	.003	.003	.002
	std	0.03	0.01	.006	.002	.0008	.001
	dim	2	2.9	3.5	4.2	4.1	4.8
$b_3$	MISE	1.115	0.596	0.144	0.123	0.046	0.043
	std	0.22	0.17	0.06	0.06	0.01	0.01
	dim	3	3.7	5.6	5.9	8.1	8.4

TABLE 2. Results for the two cases for basis T, 200 repetitions.  $100 \times \text{MISE}$  is computed from (30) as well as  $100 \times \text{std}$ . dim is the mean of selected dimensions.

Function		$N = 50$		$N = 200$		$N = 1000$	
		Case 1	Case 2	Case 1	Case 2	Case 1	Case 2
$b_1$	MISE	3.577	51.6	0.749	10.6	0.044	2.47
	std	0.09	37.9	0.02	5.07	0.003	0.54
	dim	5.0	4.0	6.0	5.0	7.0	6.0
$b_2$	MISE	0.046	0.002	.0015	.0009	.0002	.0001
	std	0.58	0.001	.0009	.0007	.0004	.0001
	dim	2.0	2.9	2.1	2.4	2.9	3.1
$b_3$	MISE	0.949	0.109	0.044	0.141	0.022	0.013
	std	0.31	0.10	0.04	0.05	0.015	0.006
	dim	3.0	4.0	4.0	6.0	4.0	6.4

TABLE 3. Results for Cases 1 and 2 for basis L, 200 repetitions.  $100 \times \text{MISE}$  is computed from (30) as well as  $100 \times \text{std}$ . dim is the mean of selected dimensions.

To test the stability of the estimation with respect to the type of the covariates, we make a permutation of the explanatory variables and consider as "Case 2" the same model and noise with leading part  $b_1 X_2 + b_2 X_3 + b_3 X_1$ . In comparison, we design as "Case 1" the initial  $b_1 X_1 + b_2 X_2 + b_3 X_3$ . The signal to noise ratio is computed here as the standard deviation of the observed  $Y$  divided by the standard deviation of the noise, and ranges between 2 and 3.

Figures 2 and 3 present 25 estimated curves compared to the true ones for the three functions, for  $N = 200$ , in Case 1 for Figure 2 and Case 2 for Figure 3. The main problem when permuting the explanatory is for the estimation of  $b_1$  in basis L for Case 2: here,  $b_1$  is associated to the Poisson explanatory variable, this seems to increase the error. This is illustrated by the first plot on the second line of Figure 3: the general form of the curve is well reconstructed, but a slight bias appears. Globally, the method works very well for these regular functions.

Tables 2 and 3 present numerical results for different sample sizes  $N = 50, 200$  and  $1000$ , in cases 1 and 2 for the basis T (Table 2) and basis L (Table 3), and for  $J = 200$  repetitions. As expected, the error decreases when  $N$  increases; moreover, the selected dimensions have not the same order for each function (the selection is adaptive to anisotropy) and increase with  $N$ . Basis L gives good results, except in the case mentioned above, but basis T seems more relevant in this problem and more stable when permuting explanatory variables. However, comparing the 18 scores in term of MISEs, L wins 10 times and T 8 times.

Clearly, many additional experiments may be driven, by changing the functions  $b_j$ , the noise level  $c_\varepsilon$ , or the noise type (Brownian type as in the lower bound Corollary 1). We do not pretend to be exhaustive, but the method is simple to implement, performs rapidly and gives globally very good results.

**5.4. Real data.** We are examining the electricity usage patterns of household appliances in a low-energy house situated in Stambrugde, Belgium. The home is equipped with sensors that measure temperature and humidity at various locations indoors and outdoors. The dataset under investigation is publicly accessible via the UCI Machine Learning Repository (<https://archive.ics.uci.edu/>). It encompasses recordings of 18 variables every 10 minutes, spanning from January 11th, 2016, at 5 pm to May 27th, 2016, at 6 pm. This dataset forms a multidimensional time series, with each dimension corresponding to one of the 18 variables. The original dataset was completed with other weather parameters from a nearby weather station that we do not consider here.

The focus is on the appliances energy consumption (in Wh) plotted in Figure 4, as it represents the main contributor to overall energy usage. We transform this time series into a functional sample by segmenting it on a daily basis and take it as the functional response variable  $Y_i(t)$  where  $i = 1, \dots, N$  ( $N = 137$  days) and  $t = 1, \dots, 144$  discretized measurements over each day (24 hours), see Figure 5 (left).

This dataset has been first analyzed by Candanedo *et al* (2017),[5]. Chagny *et al.* (2022), [6] investigated this data and seek to predict the logarithm of the energy consumption  $Y_i(t)$  for a given day  $i$  by choosing as a predictor the prior day's energy consumption  $Y_{i-1}(t)$ , resulting in an autoregressive model of order 1. Our approach is different since we want to explore the explicative functional variables available in the dataset. Following the statistical analysis of Candanedo *et al.* (2017),[5], we choose to apply our model by taking two explicative functional variables. First, the discrete variable "lights" giving the energy consumption of lights in the house has been shown to be a good predictor of room occupancy and it is the most correlated with the energy consumptions of appliances. We also choose the temperature outside the house  $T_6$  which is one of the 18 continuous variables in the dataset giving conditions in the house. Pairs plot in Figure 5 (right) show the relationship between the energy consumption of appliances with lights and temperature outside  $T_6$ . The correlation between the appliances energy consumption and both lights and  $T_6$  are respectively 0.20 and 0.12. The lower the lighting consumption, the lower the consumption of appliances decreases. The appliances consumption profile is highly variable as seen in Figure 5 with periods of almost constant demand followed by high spikes.

We kept 19,728 out of 19,735 observations in order to have 137 full days with 144 measures per day. The interval  $[0, \tau]$  with  $\tau = 1$  stands for one day and is discretized by taking  $\Delta = 1/144$ .

We also add a constant level function  $X_1(t) = 100$  in the model. We split the data into the first 136 days to estimate the coefficients and we kept the data of the 137th day to predict the appliances energy consumption.

The estimated coefficient functions of the model  $b_1, b_2, b_3$  associated with  $X_1(t) = 100$ ,  $X_2(t)$  Temperature outside and  $X_3(t)$  lights over the first 136 days are given in Figure 6. The coefficient functions  $b_2$  and  $b_3$  (respectively middle and right) are in accordance with the expected shape. The coefficient peaks correspond to times of the day when the occupants of the house are notably present in the evening and in the morning.

We also compute for each basis  $B$ , the global vector of residuals of size  $19,728 \times 1$ :

$$\Delta Y(B) := \text{vec} \left( (Y_i(t_j) - \sum_{k=1}^3 b_k(t_j) X_{i,k}(t_j))_{1 \leq i \leq 136, 1 \leq j \leq 144} \right).$$



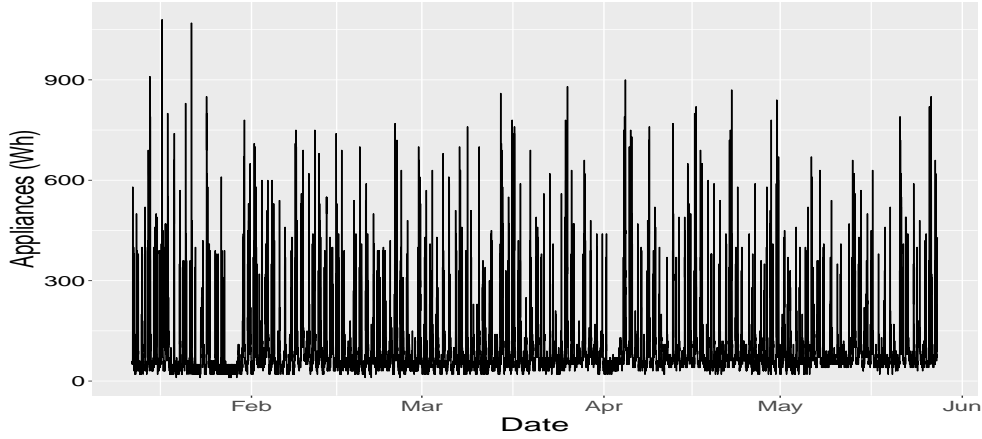


FIGURE 4. Appliances energy consumption measurements (in Wh) from the 11th January 2016 at 5 pm to the 27th May 2016 at 6 pm.

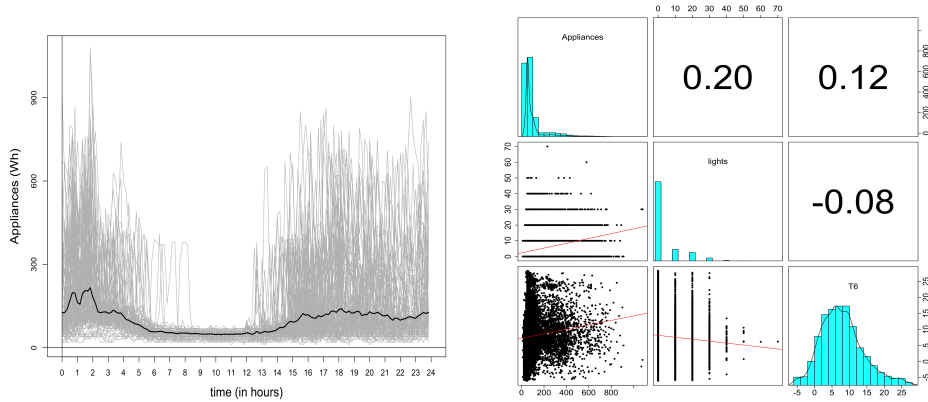


FIGURE 5. Appliances energy consumption data. Left : functional measurement data per day ( $N = 137$  curves, mean curve in bold). Right : pairs plot showing relationships between the energy consumption of appliances (response  $Y$ ) with the temperature outside  $T_6$  ( $X_2$ ) and lights ( $X_3$ ).

For the bases  $B=H, L$  or  $T$  (Hermite, Laguerre or Trigonometric), we find the following empirical residual mean and variances:  $\overline{\Delta Y(H)} = -2.325 \cdot 10^{-7}$ ,  $\overline{\Delta Y(L)} = -3.559 \cdot 10^{-9}$ ,  $\overline{\Delta Y(T)} = -2.38 \cdot 10^{-14}$  and  $\text{Var}(\Delta Y(H)) = 8.6 \cdot 10^3$ ,  $\text{Var}(\Delta Y(L)) = 8.6 \cdot 10^3$ ,  $\text{Var}(\Delta Y(T)) = 8.8 \cdot 10^3$ .

The large variance of the residuals is related to large  $\sigma$ , see Proposition 4 and here it is clear that the high spikes in the energy consumption result in high value of  $\sigma$ . The same phenomena has been observed in the prediction made by Chagny *et al.* (2022) [6] with their autoregressive model.

Lastly, we compare in Figure 7 the mean curve of  $Y$  over the 136 days of estimation to the prediction using the 144 values of the  $X_{137,k}(t_j)$ ,  $j = 1, \dots, 144$  on day 137. We can see that the global shape of the curves are not so far. Our forecast captures the consumption trend quite well, but do not allow us to detect the brutal high spikes, which is not surprising. Even estimating  $\sigma$  as a function, would not necessarily help for such prediction.

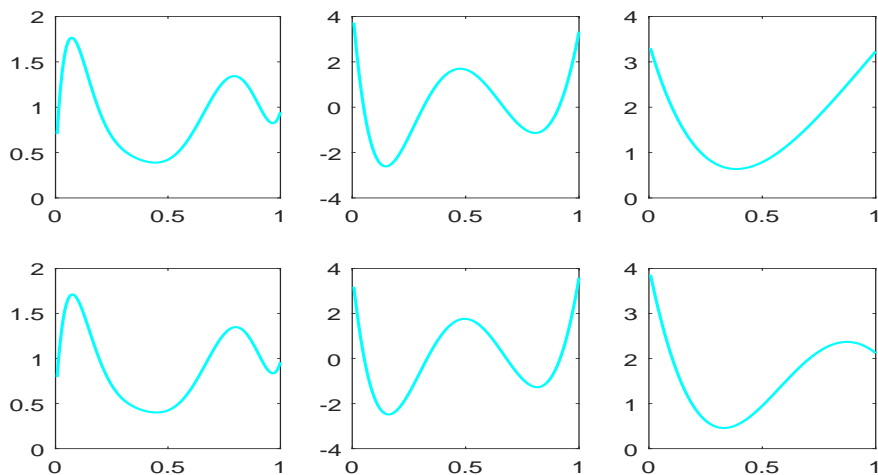


FIGURE 6. Electricity data. Estimation of  $b_1$  left,  $b_2$  middle and  $b_3$  right, with basis L (first line) and basis H (second line). Selected dimensions are  $(8,5,5)$  and  $(8,5,4)$ .

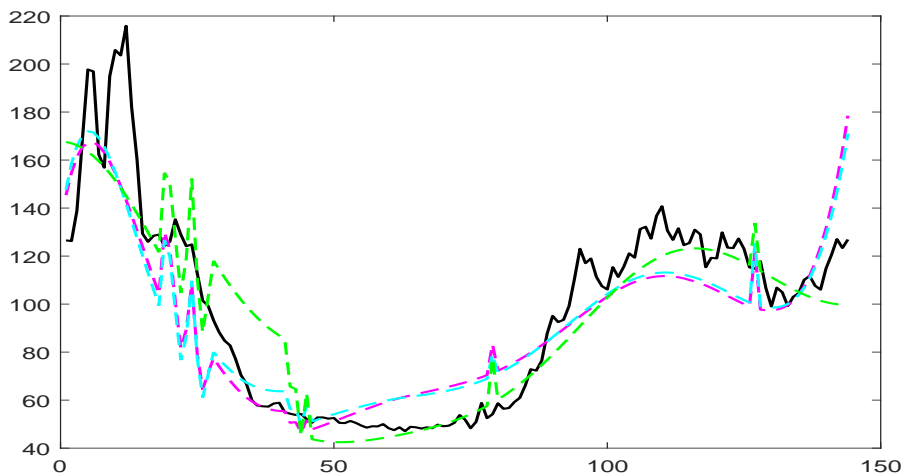


FIGURE 7. Comparison of the mean (black curve) of electricity appliance over 136 days to the prediction on the 137th day: with Hermite (magenta), Laguerre (cyan) and Trigonometric (green) estimators of  $b_j$ ,  $j = 1, 2, 3$ .

## 6. PROOFS

6.1. **Proof of Theorem 1.** We start with some preliminaries.

6.1.1. *General orthogonal projection w.r.t.  $\langle \cdot, \cdot \rangle_N$ .* To study the risk of  $\tilde{\mathbf{b}}_m$ , we need to have an adequate expression of the orthogonal projection of  $\mathbf{b}$  with respect to  $\langle \cdot, \cdot \rangle_N$ . Let

$$\Phi_{m_1+\dots+m_{j-1}+k} = \underbrace{(0, \dots, 0}_{j-1}, \varphi_k, \underbrace{0, \dots, 0}_{K-j})^T, \quad k = 1, \dots, m_j, \quad j = 1, \dots, K.$$

The functions  $(\Phi_j, j = 1, \dots, |\mathbf{m}|)$  constitute an orthonormal system of  $(\mathbb{L}_\tau^2)^K$  with respect to the scalar product  $\langle \mathbf{h}, \mathbf{h}^* \rangle = \int_0^\tau \sum_{j=1}^K h_j(t) h_j^*(t) dt$  and generate a space  $\mathbf{S}_{|\mathbf{m}|}$  (isomorphic to  $\mathbf{S}_{\mathbf{m}}$ ) with dimension  $|\mathbf{m}| = m_1 + \dots + m_K$ . An element  $\mathbf{h} = (h_1, \dots, h_K)^T$  of  $\mathbf{S}_{|\mathbf{m}|}$  can be written as

$$\mathbf{h}(t) = \sum_{i=1}^{|\mathbf{m}|} a_i \Phi_i = \left( \sum_{i=1}^{m_1} a_i \varphi_i, \sum_{i=1}^{m_2} a_{m_1+i} \varphi_i, \dots, \sum_{i=1}^{m_K} a_{m_1+\dots+m_{K-1}+i} \varphi_i \right)^T.$$

We have:

$$\widehat{\Psi}_{\mathbf{m}} = (\langle \Phi_j, \Phi_\ell \rangle_N)_{0 \leq j, \ell \leq |\mathbf{m}|}.$$

Indeed, if, for  $1 \leq j \leq m_k$  and  $1 \leq \ell \leq m_{k'}$ ,  $m_1 + \dots + m_{k-1} + j \leq m_1 + \dots + m_k$  and  $m_1 + \dots + m_{k'-1} + \ell \leq m_1 + \dots + m_{k'}$ ,  $\langle \Phi_j, \Phi_\ell \rangle_N = \widehat{\Psi}_{m_k m_{k'}}$ .

The orthogonal projection  $\pi_{\mathbf{m}} \mathbf{b}$  of  $\mathbf{b}$  on  $\mathbf{S}_{|\mathbf{m}|}$  with respect to the scalar product  $\langle \cdot, \cdot \rangle_N$  is characterized by  $\pi_{\mathbf{m}} \mathbf{b} - \mathbf{b} \perp \Phi_j, j = 1, \dots, |\mathbf{m}|$ . This yields

$$(31) \quad \pi_{\mathbf{m}} \mathbf{b} = \sum_{j=1}^{|\mathbf{m}|} \beta_j \Phi_j \quad \text{where} \quad (\beta_j)_{1 \leq j \leq |\mathbf{m}|} = \widehat{\Psi}_{\mathbf{m}}^{-1} (\langle \mathbf{b}, \Phi_j \rangle_N)_{1 \leq j \leq |\mathbf{m}|}$$

The vector  $\widehat{V}_{\mathbf{m}} = (\widehat{V}_{1, m_1}^T, \dots, \widehat{V}_{K, m_K}^T)^T$  can be written as

$$(32) \quad \widehat{V}_{\mathbf{m}} = (\langle \mathbf{b}, \Phi_j \rangle_N)_{0 \leq j \leq |\mathbf{m}|} + \mathbb{W}_{\mathbf{m}}, \quad \mathbb{W}_{\mathbf{m}} := \frac{1}{N} \left( \int_0^\tau \Phi_j(t)^T \mathbf{M}_N(t) dt \right)_{0 \leq j \leq |\mathbf{m}|}.$$

where  $\mathbf{M}_N(t) = \left( \sum_{i=1}^N X_{i,j}(t) \sigma(t, \mathbf{X}_i(t)) \varepsilon_i(t) \right)_{1 \leq j \leq K}$ . Note that, recalling the definition of  $\Theta_{\mathbf{m}}$  given in (8), we have

$$(33) \quad \mathbb{E}[\mathbb{W}_{\mathbf{m}} \mathbb{W}_{\mathbf{m}}^T] = \frac{1}{N} \Theta_{\mathbf{m}},$$

The matrix  $\Theta_{\mathbf{m}}$  is symmetric and nonnegative as, using  $h_j(t) = \sum_{p=1}^{m_j} x_{j,p} \varphi_p(t)$ , we get that  $\mathbf{x}^T \Theta_{\mathbf{m}} \mathbf{x} = \mathbb{E} \left[ \int_0^\tau \sum_{j=1}^K h_j(t) X_{1,j}(t) \sigma(t, \mathbf{X}_1(t)) \varepsilon_1(t) dt \right]^2 \geq 0$ .

We state the following Lemma, proved in Comte and Genon-Catalot (2024, Lemma 6.2) [8].

**Lemma 1.** *Assume  $(\mathcal{A}_{X,p})$  and  $(\mathcal{A}_S)$ . Define the set*

$$(34) \quad \Omega_{\mathbf{m}} := \left\{ \left| \frac{\|\mathbf{h}\|_N^2}{\|\mathbf{h}\|_\Gamma^2} - 1 \right| \leq \frac{1}{2}, \forall \mathbf{h} \in \mathbf{S}_{\mathbf{m}} \right\}.$$

where the empirical norm  $\|\cdot\|_N$  and the  $\|\cdot\|_\Gamma$ -norm are equivalent for elements of  $\mathbf{S}_{\mathbf{m}}$ . We have  $\mathcal{O}_N \subset \Omega_{\mathbf{m}}$  for all  $\mathbf{m}$ , and

$$(35) \quad \Omega_{\mathbf{m}} = \left\{ \|\Psi_{\mathbf{m}}^{-1/2} \widehat{\Psi}_{\mathbf{m}} \Psi_{\mathbf{m}}^{-1/2} - \text{Id}_{|\mathbf{m}|}\|_{\text{op}} \leq 1/2 \right\}.$$

We also recall the following proposition, proved in Comte and Genon-Catalot (2024, [8], Proposition 3.4), which shows that  $\Lambda_N$  has large probability and guarantees a rough bound on  $\|\widehat{\Psi}_{\mathbf{m}}^{-1}\|_{\text{op}}$ .

**Proposition 6.** *Assume that  $(\mathcal{A}_S)$ ,  $(\mathcal{A}_{X,p})(a)$  for  $p \geq 2$  are fulfilled. Then, there exists a constant  $c_0 > 0$  depending on  $K, \mathfrak{f}_\Gamma$  (see (12)) and  $p$ , such that  $\mathbb{P}(\Lambda_N^c) \leq c_0 N^{-p/2}$ . Moreover, on  $\Lambda_N$ , it holds that  $\forall \mathbf{m}, \quad \|\widehat{\Psi}_{\mathbf{m}}^{-1}\|_{\text{op}} \leq \mathbf{c}_1 N^{c_2}$ .*

We prove Inequality (16), proof of Inequality (17) is deferred to the Supplementary Material.

6.1.2. *Proof of inequality (16).* We write,

$$(36) \quad \begin{aligned} \|\tilde{\mathbf{b}}_{\mathbf{m}} - \mathbf{b}\|_N^2 &= \|\hat{\mathbf{b}}_{\mathbf{m}} - \mathbf{b}\|_N^2 \mathbf{1}_{\Lambda_N} + \|\mathbf{b}\|_N^2 \mathbf{1}_{\Lambda_N^c} \\ &= \|\hat{\mathbf{b}}_{\mathbf{m}} - \mathbf{b}\|_N^2 \mathbf{1}_{\Lambda_N \cap \mathcal{O}_N} + \|\hat{\mathbf{b}}_{\mathbf{m}} - \mathbf{b}\|_N^2 \mathbf{1}_{\Lambda_N \cap \mathcal{O}_N^c} + \|\mathbf{b}\|_N^2 \mathbf{1}_{\Lambda_N^c} := T_1 + T_2 + T_3. \end{aligned}$$

• Consider the term  $T_3 = \|\mathbf{b}\|_N^2 \mathbf{1}_{\Lambda_N^c}$  where  $\|\mathbf{b}\|_N^2 = \frac{1}{N} \sum_{i=1}^N \int_0^\tau \left( \sum_{j=1}^K b_j(t) X_{i,j}(t) \right)^2 dt$ . We have  $\mathbb{E}[T_3] \leq \mathbb{E}^{1/2}[\|\mathbf{b}\|_N^4] \mathbb{P}^{1/2}(\Lambda_N^c)$ . Thus, it holds that

$$\begin{aligned} \mathbb{E}[\|\mathbf{b}\|_N^4] &\leq \tau \int_0^\tau \left( \sum_{j=1}^K b_j^2(t) \right)^2 \mathbb{E} \left[ \left( \sum_{j=1}^K X_{1,j}^2(t) \right)^2 \right] dt \\ &\leq K\tau \int_0^\tau \left( \sum_{j=1}^K b_j^2(t) \right)^2 dt \sum_{j=1}^K \mathbb{E} \left[ \sup_{t \in [0, \tau]} X_{1,j}^4(t) \right] := c_K(\tau). \end{aligned}$$

Then, Proposition 6 implies  $\mathbb{E}[T_3] \leq CN^{-p_0/2} \leq \frac{C}{N}$  for  $p_0 \geq 2$ , *i.e.*  $p = 2p_0 \geq 4$ .

• Let us now study of  $T_1 = \|\hat{\mathbf{b}}_{\mathbf{m}} - \mathbf{b}\|_N^2 \mathbf{1}_{\Lambda_N \cap \mathcal{O}_N}$ . We can write:

$$(37) \quad \|\hat{\mathbf{b}}_{\mathbf{m}} - \mathbf{b}\|_N^2 = \|\hat{\mathbf{b}}_{\mathbf{m}} - \pi_{\mathbf{m}} \mathbf{b}\|_N^2 + \|\pi_{\mathbf{m}} \mathbf{b} - \mathbf{b}\|_N^2 = \|\hat{\mathbf{b}}_{\mathbf{m}} - \pi_{\mathbf{m}} \mathbf{b}\|_N^2 + \inf_{h \in \mathbf{S}_{\mathbf{m}}} \|\mathbf{b} - h\|_N^2.$$

On the one hand, we have  $\hat{\mathbf{b}}_{\mathbf{m}} = \sum_{j=1}^{|\mathbf{m}|} [\hat{\mathbf{B}}_{\mathbf{m}}]_j \Phi_j$  with  $\hat{\mathbf{B}}_{\mathbf{m}}^T = (\hat{\beta}_{1,1}, \dots, \hat{\beta}_{1,m_1}, \dots, \hat{\beta}_{K,m_K}) = \hat{\Psi}_{\mathbf{m}}^{-1} \hat{\mathbf{V}}_{\mathbf{m}}$ . On the other hand,  $\pi_{\mathbf{m}} \mathbf{b} = \sum_{j=1}^{|\mathbf{m}|} \beta_j \Phi_j$  where (see (31))  $B_{\mathbf{m}} = (\beta_1, \dots, \beta_{|\mathbf{m}|})^T = \hat{\Psi}_{\mathbf{m}}^{-1} (\langle \Phi_j, b \rangle_N)_{1 \leq j \leq |\mathbf{m}|}$ .

Hence, by (32),  $\hat{\mathbf{B}}_{\mathbf{m}} - B_{\mathbf{m}} = \hat{\Psi}_{\mathbf{m}}^{-1} \mathbb{W}_{\mathbf{m}}$  and using that if  $h = \sum_{j=1}^{|\mathbf{m}|} x_j \Phi_j$  then  $\|h\|_N^2 = \mathbf{x}^T \hat{\Psi}_{\mathbf{m}} \mathbf{x}$  with our standard notation,

$$(38) \quad \|\hat{\mathbf{b}}_{\mathbf{m}} - \pi_{\mathbf{m}} \mathbf{b}\|_N^2 = (\mathbb{W}_{\mathbf{m}})^T \hat{\Psi}_{\mathbf{m}}^{-1} \hat{\Psi}_{\mathbf{m}} \hat{\Psi}_{\mathbf{m}}^{-1} \mathbb{W}_{\mathbf{m}} = (\mathbb{W}_{\mathbf{m}})^T \hat{\Psi}_{\mathbf{m}}^{-1} \mathbb{W}_{\mathbf{m}}.$$

Recall that by Lemma 1,  $\mathcal{O}_N \subset \Omega_{\mathbf{m}}$ . On  $\Omega_{\mathbf{m}} = \left\{ \|\Psi_{\mathbf{m}}^{-1/2} \hat{\Psi}_{\mathbf{m}} \Psi_{\mathbf{m}}^{-1/2} - \text{Id}_{|\mathbf{m}|}\|_{\text{op}} \leq 1/2 \right\}$ , all the eigenvalues of  $\Psi_{\mathbf{m}}^{-1/2} \hat{\Psi}_{\mathbf{m}} \Psi_{\mathbf{m}}^{-1/2}$  belong to  $[1/2, 3/2]$  and so all the eigenvalues of  $\Psi_{\mathbf{m}}^{1/2} \hat{\Psi}_{\mathbf{m}}^{-1} \Psi_{\mathbf{m}}^{1/2}$  belong to  $[2/3, 2]$ . Thus, we write

$$(39) \quad \begin{aligned} (\mathbb{W}_{\mathbf{m}})^T \hat{\Psi}_{\mathbf{m}}^{-1} \mathbb{W}_{\mathbf{m}} \mathbf{1}_{\mathcal{O}_N} &= (\mathbb{W}_{\mathbf{m}})^T \Psi_{\mathbf{m}}^{-1/2} \Psi_{\mathbf{m}}^{1/2} \hat{\Psi}_{\mathbf{m}}^{-1} \Psi_{\mathbf{m}}^{1/2} \Psi_{\mathbf{m}}^{-1/2} \mathbb{W}_{\mathbf{m}} \mathbf{1}_{\mathcal{O}_N} \\ &\leq 2(\mathbb{W}_{\mathbf{m}})^T \Psi_{\mathbf{m}}^{-1} \mathbb{W}_{\mathbf{m}} \mathbf{1}_{\mathcal{O}_N}. \end{aligned}$$

Therefore, by using equality (33), and in particular  $\mathbb{E}[[\mathbb{W}_{\mathbf{m}}]_j [\mathbb{W}_{\mathbf{m}}]_k] = [\Theta_{\mathbf{m}}]_{j,k}$ , we get

$$(40) \quad \begin{aligned} \mathbb{E} \left[ \|\hat{\mathbf{b}}_{\mathbf{m}} - \pi_{\mathbf{m}} \mathbf{b}\|_N^2 \mathbf{1}_{\mathcal{O}_N \cap \Lambda_N} \right] &\leq 2\mathbb{E} \left[ \sum_{1 \leq j, k \leq \mathbf{M}} [\mathbb{W}_{\mathbf{m}}]_j [\mathbb{W}_{\mathbf{m}}]_k [\Psi_{\mathbf{m}}^{-1}]_{j,k} \right] \\ &= \frac{2}{N} \sum_{1 \leq j, k \leq \mathbf{M}} [\Psi_{\mathbf{m}}^{-1}]_{j,k} [\Theta_{\mathbf{m}}]_{j,k} = \frac{2}{N} \text{Tr}[\Psi_{\mathbf{m}}^{-1} \Theta_{\mathbf{m}}], \end{aligned}$$

So we obtain:

$$\mathbb{E}[T_1] \leq \mathbb{E} \left[ \inf_{h \in \mathbf{S}_{\mathbf{m}}} \|\mathbf{b} - h\|_N^2 \right] + \frac{2}{N} \text{Tr}[\Psi_{\mathbf{m}}^{-1} \Theta_{\mathbf{m}}] \leq \inf_{h \in \mathbf{S}_{\mathbf{m}}} \|\mathbf{b} - h\|_N^2 + \frac{2}{N} \text{Tr}[\Psi_{\mathbf{m}}^{-1} \Theta_{\mathbf{m}}],$$

where the second term of the right-hand-side (rhs) above is the variance term appearing in (16).

• Finally, let us study of  $T_2 = \|\widehat{\mathbf{b}}_{\mathbf{m}} - \mathbf{b}\|_N^2 \mathbf{1}_{\Lambda_N \cap \mathcal{O}_N^c}$ . We have  $T_2 \leq (\|\widehat{\mathbf{b}}_{\mathbf{m}} - \pi_{\mathbf{m}} \mathbf{b}\|_N^2 + \|\mathbf{b}\|_N^2) \mathbf{1}_{\Lambda_N \cap \mathcal{O}_N^c}$ . Using (38) yields

$$(41) \quad T_2 \leq (\mathbb{W}_{\mathbf{m}}^T \widehat{\Psi}_{\mathbf{m}}^{-1} \mathbb{W}_{\mathbf{m}} + \|\mathbf{b}\|_N^2) \mathbf{1}_{\Lambda_N \cap \mathcal{O}_N^c}.$$

By Proposition 6 about  $\Lambda_N$  and the Cauchy-Schwarz inequality, we get,

$$(42) \quad \mathbb{E}[T_2] \leq \left( 2c_1 N^{c_2} \mathbb{E}^{1/2}[(\mathbb{W}_{\mathbf{m}}^T \mathbb{W}_{\mathbf{m}})^2] + \mathbb{E}^{1/2}[\|\mathbf{b}\|_N^4] \right) \mathbb{P}^{1/2}(\mathcal{O}_N^c).$$

We have already seen that  $\mathbb{E}[\|\mathbf{b}\|_N^4] \leq c_K(\tau)$ . For the term  $\mathbb{E}[(\mathbb{W}_{\mathbf{m}}^T \mathbb{W}_{\mathbf{m}})^2]$ , we prove the following:

**Lemma 2.** *Under the Assumption of Theorem 1, with  $\mathbb{W}_{\mathbf{m}}$  defined in (32), we have, for some constant  $c^* := c(\tau, \mathfrak{M}_\sigma, c_\varepsilon, c_\varphi)$ , that  $\mathbb{E}[(\mathbb{W}_{\mathbf{m}}^T \mathbb{W}_{\mathbf{m}})^2] \leq c^*(1 \vee N^{\omega-1})$ .*

Plugging the result of Lemma 2 in (42) allows to conclude, with the bound in Proposition 3, that, for all  $\mathbf{m}$  satisfying  $|\mathbf{m}| \leq N$ ,  $\mathbb{E}[T_2] \leq CN^{c_2-p/2+(\omega-1)+/2} \leq CN^{-1}$ , for  $p \geq 2c_2 + 1 + \omega \vee 1$ . Joining the bounds for the expectations of  $T_1, T_2, T_3$  gives Inequality (16) by choosing  $p \geq 2c_2 + 1 + \omega \vee 1$  and  $p \geq 4$ .

6.1.3. *Bound on  $\text{Tr}[\Psi_{\mathbf{m}}^{-1} \Theta_{\mathbf{m}}]$ .* As  $\Psi_{\mathbf{m}}^{-1}$  and  $\Theta_{\mathbf{m}}$  are symmetric and nonnegative, we apply:

**Lemma 3.** *Let  $A, B$  be two symmetric nonnegative  $d \times d$  matrices. Then,  $\text{Tr}(AB) \leq \|A\|_{\text{op}} \text{Tr}(B)$ .*

We get  $\text{Tr}[\Psi_{\mathbf{m}}^{-1} \Theta_{\mathbf{m}}] \leq \|\Psi_{\mathbf{m}}^{-1}\|_{\text{op}} \text{Tr}[\Theta_{\mathbf{m}}] \leq \mathfrak{f}_\Gamma \sum_{j=1}^K \text{Tr}[\Theta_{m_j m_j}]$  where

$$\begin{aligned} \text{Tr}[\Theta_{m_j m_j}] &= \sum_{p=1}^{m_j} \int_0^\tau \int_0^\tau \varphi_p(t) \varphi_p(s) \mathbb{E}[X_{1,j}(t) X_{1,j}(s) \sigma(t, \mathbf{X}_1(t)) \sigma(s, \mathbf{X}_1(s))] c(s, t) ds dt \\ &\leq \mathbf{c} \sum_{j=1}^{m_j} \left( \int_0^\tau |\varphi_p(t)| \mathbb{E}^{1/2}[X_{1,j}^2(t) \sigma^2(t, \mathbf{X}_1(t))] dt \right)^2 \\ &\leq \mathbf{c}_\varepsilon \tau (m_j \wedge \tau L(S_{m_j})) \sup_{1 \leq j \leq K} \sup_{t \in [0, \tau]} \mathbb{E}[X_{1,j}^2(t) \sigma^2(t, \mathbf{X}_1(t))]. \end{aligned}$$

Therefore, using Assumption  $(A_\varphi)$ ,

$$\text{Tr}[\Psi_{\mathbf{m}}^{-1} \Theta_{\mathbf{m}}] \leq \mathfrak{f}_\Gamma \mathbf{c} \sup_{j=1, \dots, K} \sup_{t \in [0, \tau]} \mathbb{E}[X_{1,j}^2(t) \sigma^2(t, \mathbf{X}_1(t))] \left( |\mathbf{m}| \wedge (c_\varphi \tau \sum_{j=1}^K m_j^\omega) \right).$$

This leads to Inequality (18), using (15). Inequality (17) is given in Supplementary Material, to complete the proof of Theorem 1.  $\square$

6.2. **Proof of Theorem 2.** Without loss of generality suppose that  $s_1 \leq s_k$  for all  $k \in \{2, \dots, K\}$ . First for all  $\mathbf{b}$  and  $\widehat{\mathbf{b}}$  it holds, using Proposition 2, that

$$\|\widehat{\mathbf{b}} - \mathbf{b}\|_\Gamma^2 \geq \frac{1}{\mathfrak{f}_\Gamma} \|\widehat{\mathbf{b}} - \mathbf{b}\|^2 \geq \frac{1}{\mathfrak{f}_\Gamma} \|\widehat{b}_1 - b_1\|^2 \geq \frac{1}{\mathfrak{f}_\Gamma G_{21}^2} \|\widehat{b}_1 - b_1\|_{\Gamma_1}^2,$$

where  $G_{21}^2 := \sup_{t \in [0, \tau]} \mathbb{E}[X_{11}^2(t)]$  and  $\|u\|_{\Gamma_1}^2 = \int_0^\tau u^2(t) \mathbb{E}[X_1^2(t)] dt$ . Therefore, to establish the result, it is enough to prove that, for  $C > 0$ ,

$$(43) \quad \inf_{\widehat{b}_1} \sup_{b_1 \in W(s_1, L_1)} \mathbb{E}[\|\widehat{b}_1 - b_1\|_{\Gamma_1}^2] \geq CN^{-2s_1/(2s_1+1)}.$$

In the sequel for sake of readability we drop the dependency in the index 1 for the function  $b$  and the regularity  $s$ .

Following Theorem 2.7 in Tsybakov (2009) [27], the lower bound is established by considering a decision problem between an increasing number  $Q+1$  of competing functions  $b_0, \dots, b_Q$  contained in  $W(s, L)$  for some positive constants  $s, L$ . The following two conditions must be satisfied

- (i)  $\|b_q - b_{q'}\|_{\Gamma_1}^2 \geq dN^{-\frac{2s}{2s+1}}, \forall 0 \leq q < q' \leq Q$  for some positive constant  $d$ ,
- (ii)  $\frac{1}{Q} \sum_{q=1}^Q \text{KL}(\mathbb{P}_q^{\otimes N}, \mathbb{P}_0^{\otimes N}) \leq \kappa \log Q$ , for some  $0 < \kappa < 1/8$ ,

where  $\mathbb{P}_q$  denotes the distribution of  $(X(t), Y^q(t))_{t \in [0, \tau]}$ , where  $Y^q(t) := b_q(t)X(t) + \sigma(t)\varepsilon(t)$  and KL is the Kullback Leibler divergence.

Consider  $b_0 = 0$  which belongs to the class  $W(s, L)$  and the functions  $b_q = \delta M^{-\gamma} \sum_{j=1}^M \theta_{q,j} \varphi_j$ , for some  $\delta > 0, \gamma > 0$  and  $M > 0$  that will be specified below and  $\theta_q = (\theta_{q,1}, \dots, \theta_{q,M}) \in \{0, 1\}^M$ . To ensure that the functions  $b_q$  belong to  $W(s, L)$  for  $1 \leq q \leq Q$  we choose  $(\delta, \gamma, M)$  such that

$$\delta^2 M^{-2\gamma} \sum_{j=2}^M j^{2s} \theta_{q,j}^2 \leq \delta^2 M^{-2\gamma} \sum_{j=2}^M j^{2s} \leq \delta^2 M^{2s+1-2\gamma} \leq L^2, \text{ for } \gamma \geq s + \frac{1}{2} \text{ and } 0 < \delta \leq L.$$

Note that this implies that  $\delta^2 M^{-2\gamma} \sum_{J \leq j \leq M} \theta_{q,j}^2 \leq L^2 J^{-2s}$  and thus  $b_q \in W(s, L)$ .

Next we show a Lemma that implies point (i) above.

**Lemma 4.** *Suppose  $(A_S)$  and let  $M \geq 8$ , there exist  $Q \geq 2^{M/8}$  elements  $\{\theta_1, \dots, \theta_Q\}$  of  $\{0, 1\}^M$ ,  $\theta_0 = (0, \dots, 0)$ , such that it holds  $\|b_q - b_{q'}\|_{\Gamma_1}^2 \geq \frac{1}{8} \frac{\delta^2}{f_{\Gamma}} M^{-2\gamma+1}, \forall 0 \leq q < q' \leq Q$ .*

*Proof of Lemma 4.* Let  $\theta_q, \theta_{q'} \in \{0, 1\}^Q$ , we compute

$$\|b_q - b_{q'}\|_{\Gamma_1}^2 = \int_0^{\tau} \delta^2 M^{-2\gamma} \left( \sum_{j=1}^M (\theta_{q,j} - \theta'_{q,j}) \varphi_j(x) \right)^2 \mathbb{E}[X(t)^2] dx \geq \frac{\delta^2}{f_{\Gamma}} M^{-2\gamma} \rho(\theta_q, \theta_{q'})$$

where we used  $(A_S)$  (for  $K = 1$  it holds  $\Gamma_1(t) = \mathbb{E}[X_{\Gamma_1}^2(t)]$ ) and where  $\rho$  denotes the Hamming distance defined by  $\rho(\theta_q, \theta_{q'}) = \sum_{j=1}^M \mathbf{1}_{\theta_{q,j} \neq \theta_{q',j}}$ . Next the Varshamov-Gilbert bound (see Lemma 2.9 in Tsybakov (2009) [27]) ensures that for  $M \geq 8$  there exist  $Q \geq 2^{M/8}$  elements  $\{\theta_0, \dots, \theta_Q\}$  of  $\{0, 1\}^M$  such that  $\rho(\theta_q, \theta_{q'}) \geq M/8$  for all  $0 \leq q < q' \leq Q$  with  $\theta_0 = (0, \dots, 0)$ . This leads to the desired result.  $\square$

Now, we show a Lemma that enables to establish point (ii) above.

**Lemma 5.** *Let  $\theta \in \{0, 1\}^M$ , suppose that  $(A_{X,p})$  holds with  $p = 2$  and  $\sigma\varepsilon$  as in Theorem 2. Then, it holds that*

$$\text{KL}(\mathbb{P}_q^{\otimes N}, \mathbb{P}_0^{\otimes N}) \leq \delta^2 c_{\sigma\varepsilon}^2 \frac{N}{2} G_2^2 M^{-2\gamma+1}.$$

*Proof of Lemma 5.* To establish this result we rely on the properties of the Kullback Leibler divergence and the Cameron Martin theorem (Chagny *et al.* [6] Theorem 4, see also Lifshits (2012) [23]). Consider the Hilbert space  $(\mathbb{L}_{\tau}^2, \langle \cdot, \cdot \rangle, \|\cdot\|)$  and  $\sigma\varepsilon$  a centered Gaussian random process with distribution  $\mathbb{P}_0$  and covariance operator  $\Sigma_{\sigma\varepsilon}$  satisfying the assumptions of the Theorem. Consider the subset  $H_P = \{h \in \mathbb{L}_{\tau}^2, \langle h, \Sigma_{\sigma\varepsilon}^{-1} h \rangle < \infty\}$ , for all  $h \in H_P$  it holds that the distribution of  $\sigma\varepsilon + h$  is absolutely continuous with respect to  $\mathbb{P}_0$  and its density is given by

$$\frac{d\mathbb{P}_h}{d\mathbb{P}_0}(x) = \exp \left( \langle x, \Sigma_{\sigma\varepsilon}^{-1} h \rangle - \frac{1}{2} \langle h, \Sigma_{\sigma\varepsilon}^{-1} h \rangle \right).$$

It allows to write for  $1 \leq q \leq Q$  and  $b_q$  using that  $b_q X \in H_P$ , conditioning on  $X$ , that

$$\begin{aligned} \text{KL}(\mathbb{P}_q^{\otimes N}, \mathbb{P}_0^{\otimes N}) &= N \text{KL}(\mathbb{P}_q, \mathbb{P}_0) = N \int \log \left( \frac{d\mathbb{P}_q}{d\mathbb{P}_0} \right) d\mathbb{P}_q = N \int \int \log \left( \frac{d\mathbb{P}_{Y^q|X}}{d\mathbb{P}_{Y^0|X}} \right) d\mathbb{P}_{Y^q|X} d\mathbb{P}_X \\ &= N \mathbb{E}_X \left[ \mathbb{E}_\varepsilon \left[ \langle Y^q, \Sigma_{\sigma_\varepsilon}^{-1} b_q X \rangle - \frac{1}{2} \|\Sigma_{\sigma_\varepsilon}^{-1/2} b_q X\|^2 \right] \right] \\ &= N \mathbb{E}_X \left[ \langle b_q X, \Sigma_{\sigma_\varepsilon}^{-1} b_q X \rangle - \frac{1}{2} \langle b_q X, \Sigma_{\sigma_\varepsilon}^{-1} b_q X \rangle \right] = \frac{N}{2} \mathbb{E}_X [\langle b_q X, \Sigma_{\sigma_\varepsilon}^{-1} b_q X \rangle] \\ &\leq \frac{N}{2} \|\Sigma_{\sigma_\varepsilon}^{-1}\|_{\text{op}} \|b_q\|_\Gamma^2 \leq \frac{N}{2} \|\Sigma_{\sigma_\varepsilon}^{-1}\|_{\text{op}} \sup_{t \in [0,1]} \mathbb{E}[X(t)^2] \|b_q\|^2, \end{aligned}$$

where we used  $(\mathcal{A}_{X,p})$  for  $p = 2$  and  $\|\Sigma_{\sigma_\varepsilon}^{-1}\|_{\text{op}} < c_{\sigma_\varepsilon}^2$ . Using that  $\|b_q\|^2 \leq \delta^2 M^{-2\gamma+1}$  we get the desired result.  $\square$

Take  $\gamma = s + \frac{1}{2}$ , which ensures that  $b_q$  belongs to  $W(s, L)$  and  $M = N^{\frac{1}{2\gamma}}$ . This leads to  $\|b_q - b_{q'}\|_\Gamma^2 \geq dN^{-\frac{2s}{2s+1}}$  with Lemma 4, for some positive constant  $d$ . Moreover, using that for  $M \geq 8$ ,  $Q \geq 2^{M/8}$  it holds that  $NM^{-2\gamma+1} = M \leq \frac{8}{\log 2} \log Q$ . Injecting this in Lemma 5 and taking  $\delta^2 < ((\log 2)(2^5 c_{\sigma_\varepsilon}^2 G_2^2)^{-1}) \wedge \frac{\pi^{2s}}{L}$ , allows to write

$$\frac{1}{Q} \sum_{q=1}^Q \text{KL}(\mathbb{P}_q^{\otimes N}, \mathbb{P}_0^{\otimes N}) \leq \kappa \log Q$$

for some  $\kappa \in (0, \frac{1}{8})$ . Applying Theorem 2.7 in Tsybakov (2009) [27] implies, for some positive constant  $c$ ,  $\inf_{\hat{b}} \sup_{b \in W(s, L)} \mathbb{E} [\|\hat{b} - b\|_\Gamma^2] \geq cN^{-\frac{2s}{2s+1}}$ , and completes the proof.  $\square$

**6.3. Proof of Corollary 1.** All we need to check is that  $\sigma_1 \varepsilon_1$  (resp.  $\sigma_2 \varepsilon_2$ ) satisfies the assumptions of Theorem 2. The fact that  $\varepsilon_1$  (resp.  $\varepsilon_2$ ) is a centered Gaussian process such that  $\mathbb{E}[\varepsilon_1^2(t)] = 1$  (resp.  $\mathbb{E}[\varepsilon_2^2(t)] = 1$ ) for all  $t \in [0, 1]$  is easily checked. Similarly observe that  $\sigma_1$  (resp.  $\sigma_2$ ) are bounded on  $[0, 1]$ . Next we identify the operators  $\Sigma_{\sigma_1 \varepsilon_1}$  and  $\Sigma_{\sigma_1 \varepsilon_1}^{-1}$  (resp.  $\Sigma_{\sigma_2 \varepsilon_2}$  and  $\Sigma_{\sigma_2 \varepsilon_2}^{-1}$ ) defined by  $\forall f, g \in L_2([0, 1])$

$$\text{Cov}(\langle f, \sigma_1 \varepsilon_1 \rangle, \langle g, \sigma_1 \varepsilon_1 \rangle) = \langle f, \Sigma_{\sigma_1 \varepsilon_1} g \rangle.$$

Simple computations lead to, for all  $t \in [0, 1]$ ,

$$\Sigma_{\sigma_1 \varepsilon_1} g(t) = \int_0^1 g(s)(s \wedge t + 1) ds, \quad (\text{resp.} \quad \Sigma_{\sigma_2 \varepsilon_2} g(t) = \int_0^1 g(s) e^{-\frac{|t-s|}{2}} ds.)$$

To compute its inverse we observe that differentiating twice in  $t$  the latter leads to

$$\begin{aligned} g &= \frac{1}{4} \Sigma_{\sigma_1 \varepsilon_1} g - (\Sigma_{\sigma_1 \varepsilon_1} g)'' = \left( \frac{1}{4} I - D_2 \right) \Sigma_{\sigma_1 \varepsilon_1} g \\ (\text{resp.} \quad g &= -(\Sigma_{\sigma_2 \varepsilon_2} g)'' = -D_2(\Sigma_{\sigma_2 \varepsilon_2} g)), \end{aligned}$$

where  $D_2 : g \mapsto g''$  and  $I$  is the identity. It follows that  $\Sigma_{\sigma_1 \varepsilon_1}^{-1} = \frac{1}{4} I - D_2$  (resp.  $\Sigma_{\sigma_2 \varepsilon_2}^{-1} = -D_2$ ). Combined with the assumptions that  $X_{1k}$  is of class  $C^2([0, 1])$  and that  $s_k \geq 2$ , implying that  $b_k \in W(s_k, L_k)$  is of class  $C^2([0, 1])$  (see Proposition 1.14 of Tsybakov (2009) [27]), it ensures

that  $\langle b_k X_{1k}, \Sigma_{\sigma_1 \varepsilon_1}^{-1} b_k X_{1k} \rangle < \infty$  (resp.  $\langle b_k X_{1k}, \Sigma_{\sigma_2 \varepsilon_2}^{-1} b_k X_{1k} \rangle < \infty$ ) for all  $b_k \in W(s_k, L_k)$ . Finally, using Definition 1.11 and Proposition 1.14 of Tsybakov (2009) [27] we get

$$\begin{aligned} \|\Sigma_{\sigma_1 \varepsilon_1}^{-1}\|_{\text{op}} &= \sup_{\|g\|=1} \|\Sigma_{\sigma_1 \varepsilon_1}^{-1} g\| \leq \frac{1}{4} + \sup_{\substack{\|g\|=1 \\ g=b_k X_{1k}, b_k \in W(s_k, L_k)}} \|g''\| \leq \frac{1}{4} + c_{L_k, X_{1k}} \\ (\text{resp. } \|\Sigma_{\sigma_2 \varepsilon_2}^{-1}\|_{\text{op}} &= \sup_{\|g\|=1} \|\Sigma_{\sigma_2 \varepsilon_2}^{-1} g\| \leq \sup_{\substack{\|g\|=1 \\ g=b_k X_{1k}, b_k \in W(s_k, L_k)}} \|g''\| \leq c_{L_k, X_{1k}}), \end{aligned}$$

where  $c_{L_k, X_{1k}} := L_k^2 (\sup_{t \in [0, \tau]} |X_{1k}(t)| + 2 \sup_{t \in [0, \tau]} |X'_{1k}(t)| + \sup_{t \in [0, \tau]} |X''_{1k}(t)|)$ . This completes the proof with  $c_{\sigma_1 \varepsilon_1} := \frac{1}{4} + \max_{1 \leq k \leq K} c_{L_k, X_{1k}}$  (resp.  $c_{\sigma_2 \varepsilon_2} = \max_{1 \leq k \leq K} c_{L_k, X_{1k}}$ ).  $\square$

**Supplementary material contains** the proofs of Propositions 1, 2, 3, 4 and 5, Inequality (17) of Theorem 1, Lemma 2 and Theorem 3.

## REFERENCES

- [1] Baraud, Y. (2002). Model selection for regression on a random design. *ESAIM Probab. Statist.*, **6**, 127-146.
- [2] Bhattacharjee, S. and Müller, H. G. (2022). Concurrent object regression. *Electron. J. Stat.*, **16**, 4031-4089.
- [3] Baudry, J.-P., Maugis, C. and Michel, B. (2012) Slope heuristics: overview and implementation. *Statistics and Computing*, **22**, 455-470.
- [4] Birgé, L., and Massart, P. (1998) Minimum contrast estimators on sieves: exponential bounds and rates of convergence. *Bernoulli* **4**, 329-375.
- [5] Candanedo, L. M. Feldheim, V. and Deramaix, D. (2017). Data driven prediction models of energy use of appliances in a low-energy house. *Energy and Buildings*, **140**, 81-97.
- [6] Chagny, G. Meynaoui, A. and Roche, A. (2022). Adaptive nonparametric estimation in the functional linear model with functional output. *Preprint hal-03579232*
- [7] Comte, F. and Genon-Catalot, V. (2018). Laguerre and Hermite bases for inverse problems. *Journal of the Korean Statistical Society*, **47**, 273-296.
- [8] Comte, F. and Genon-Catalot, V. (2024). Nonparametric estimation for i.i.d. Stochastic Differential Equations with space-time dependent coefficients. *Preprint hal-04139052*. To appear in *SIAM/ASA Journal of Uncertainty Quantification*.
- [9] Comte, F. and Lacour, C. (2023) Non compact estimation of the conditional density from direct or noisy data. *Ann. Inst. Henri Poincaré, Probab. Stat.* **59**, no. 3, 1463-1507.
- [10] Comte, F. and Marie, N. (2023). On a Projection Estimator of the Regression Function Derivative. *Journal of Nonparametric Statistics*, **35**, 773-819.
- [11] Efromovich, S. (1999). *Nonparametric curve estimation. Methods, theory, and applications*. Springer Series in Statistics. Springer-Verlag, New York.
- [12] Eubank, R. L., Huang, C., Munoz M. Y., Wang, N., Wang, S., Buchanan, R. J. (2004) Smoothing spline estimation in varying-coefficient models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **66**, no.3, 653-667.
- [13] Fan, J., Yao, Q. and Cai, Z. (2003) Adaptive varying-coefficient linear models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **65**, no.1, 57-80.
- [14] Garsia, A. M., Rodemich, E. and Rumsey, Jr, H. (1970/71) A real variable lemma and the continuity of paths of some Gaussian processes. *Indiana Univ. Math. J.*, **20**, 565-578.
- [15] Ghosal, R. and Maity, A. (2022a) Variable selection in nonparametric functional concurrent regression. *Canad. J. Statist.* **50**, no.1, 142-161.
- [16] Ghosal, R. and Maity, A. (2022b) A score based test for functional linear concurrent regression. *Econom. Stat.* **21**, 114-130.
- [17] Gloter, A. (2000) Discrete sampling of an integrated diffusion process and parameter estimation of the diffusion coefficient. *ESAIM Probab. Statist.* **4**, 205-227.
- [18] Jourdain, B. and Pagès, G. (2022). Convex ordering for stochastic Volterra equations and their Euler schemes. Preprint arXiv:2211.10186v1. To appear in *Finance and Stochastics*.
- [19] Huang, J., Wu, C. O., Zhou, L. (2002) . Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika* **89**, no.1, 111-128.
- [20] Kato, T. (1995) *Perturbation Theory for Linear Operators*. Springer-Verlag, Berlin.



- [21] Li, D., Phillips, P. C. B., Gao, J. (2020) Kernel-based inference in time-varying coefficient cointegrating regression. *J. Econometrics* **215**, no. 2, 607-632.
- [22] Li, H., Duan, X. and Yin, G. (2016) Generalized method of moments for additive hazards model with clustered dental survival data. *Scand. J. Stat.* **43**, no.4, 1124-1139.
- [23] Lifshits, M. (2012) *Lectures on Gaussian processes*. SpringerBriefs Math., Springer Heidelberg.
- [24] Manrique, T., Crambes, C. and Hilgert, N. (2018) Ridge regression for the functional concurrent model. *Electron. J. Stat.* **12**, no.1, 985-1018.
- [25] Ramsay, J.O. and Silverman, B.W. (2005). *Functional data analysis*, Springer Ser. Statist. Springer, New York, xx+426 pp.
- [26] Teodorescu, B., Van Keilegom, I. and Cao, R. (2010) Generalized time-dependent conditional linear models under left truncation and right censoring. *Ann. Inst. Statist. Math.* **62**, no.3, 465-485.
- [27] Tsybakov, A. B. (2009). *Introduction to nonparametric estimation*. Revised and extended from the 2004 French original. Translated by Vladimir Zaiats. Springer Series in Statistics. Springer, New York.
- [28] Yao, F., Müller, H.-G. and Wang, J.-L. (2005a) Functional data analysis for sparse longitudinal data. *J. Amer. Statist. Assoc.* **100**, no.470, 577-590.
- [29] Yao, F., Müller, H.-G. and Wang, J.-L. (2005b) Functional linear regression analysis for longitudinal data. *Ann. Statist.* **33**, no.6, 2873-2903.

## 7. SUPPLEMENTARY MATERIAL

**7.1. Proof of Proposition 1.** The result follows from the following equalities.

For  $\mathbf{x} = (x_{1,1}, \dots, x_{1,m_1}, x_{2,1}, \dots, x_{2,m_2}, \dots, x_{K,1}, \dots, x_{K,m_K})^T \in \mathbb{R}^{|\mathbf{m}|}$ , set for  $j = 1, \dots, K$ ,  $h_j(t) = \sum_{p=1}^{m_j} x_{j,p} \varphi_p(t)$  and  $\mathbf{h} = (h_1, \dots, h_K)^T$  (see (4)). Then we have

$$\mathbf{x}^T \widehat{\Psi}_{\mathbf{m}} \mathbf{x} = \int_0^\tau \frac{1}{N} \sum_{i=1}^N \left( \sum_{j=1}^K h_j(t) X_{i,j}(t) \right)^2 dt = \|\mathbf{h}\|_N^2,$$

and  $\mathbf{x}^T \Psi_{\mathbf{m}} \mathbf{x} = \|\mathbf{h}\|_\Gamma^2$ . Indeed, using the product of matrices by blocks and setting for  $j = 1, \dots, k$ ,  $\mathbf{x}_j^T = (x_{j,1}, \dots, x_{j,m_j})$ , we get:

$$\mathbf{x}^T \widehat{\Psi}_{\mathbf{m}} \mathbf{x} = \sum_{1 \leq j, k \leq K} \mathbf{x}_j^T \widehat{\Psi}_{m_j, m_k} \mathbf{x}_k.$$

Using the definition of  $\widehat{\Psi}_{m_j, m_k}$  yields, for  $h_j$  as defined above

$$\mathbf{x}_j^T \widehat{\Psi}_{m_j, m_k} \mathbf{x}_k = \int_0^\tau h_j(t) h_k(t) \frac{1}{N} \sum_{i=1}^N X_{i,j}(t) X_{i,k}(t) dt.$$

Thus,

$$\mathbf{x}^T \widehat{\Psi}_{\mathbf{m}} \mathbf{x} = \frac{1}{N} \sum_{i=1}^N \int_0^\tau \left( \sum_{j=1}^K h_j(t) X_{i,j}(t) \right)^2 dt = \int_0^\tau \mathbf{h}(t)^T \Gamma_N(t) \mathbf{h}(t) dt.$$

Now,  $\mathbf{x}^T \widehat{\Psi}_{\mathbf{m}} \mathbf{x} = 0$  implies that  $\mathbf{h}(t)^T \Gamma_N(t) \mathbf{h}(t) = 0$  *a.e.* on  $[0, \tau]$ , by  $(\mathcal{A}_S)$ . As, for all  $j$ , the functions  $(\varphi_j, j = 1, \dots, m_j)$  are orthonormal on  $\mathbb{L}_\tau^2$ , this implies that for all  $j$ ,  $\mathbf{x}_j = 0$ , therefore,  $\mathbf{x} = 0$ . This shows that  $\widehat{\Psi}_{\mathbf{m}}$  is positive definite. The same holds for  $\Psi_{\mathbf{m}}$ .

Moreover, the constraint of finiteness of  $\mathfrak{f}_\Gamma$  defined by (12), implies the second result, it is proved in Proposition 3.2 in Comte and Genon-Catalot (2024) [8].  $\square$

**7.2. Proof of Proposition 2.** It holds that

$$\|\mathbf{h}\|_\Gamma^2 = \int_0^\tau \mathbf{h}(t)^T \Gamma(t) \mathbf{h}(t) dt \leq \sup_{t \in [0, \tau]} \|\Gamma(t)\|_{\text{op}} \|\mathbf{h}\|^2.$$

As, under  $(\mathcal{A}_{X,p})$ ,  $\|\Gamma(t)\|_{\text{op}} \leq \text{Tr}(\Gamma(t)) = \sum_{j=1}^K \mathbb{E}[X_{1,j}^2(t)] \leq K G_2^2$ , we get the first result of Proposition 2. Under  $(\mathcal{A}_S)$  we write, using that  $\mathbf{h} = (h_1, \dots, h_K)^T$  that

$$\|\mathbf{h}\|^2 = \int_0^\tau \sum_{j=1}^K h_j^T(t) S^{1/2}(t) S^{-1}(t) S^{1/2}(t) h_j(t) dt \leq \sup_{t \in [0, \tau]} \|\Gamma(t)^{-1}\|_{\text{op}} \|\mathbf{h}\|_\Gamma^2 = \mathfrak{f}_\Gamma \|\mathbf{h}\|_\Gamma^2,$$

which completes the proof.

**7.3. Proof of Proposition 3.** Define the event:

$$(44) \quad \mathcal{O}_N = \left\{ \sup_{t \in [0, \tau]} \|\Gamma(t)^{-1/2} \Gamma_N(t) \Gamma(t)^{-1/2} - \text{Id}_K\|_{\text{op}} \leq \frac{1}{2} \right\}.$$

To get the announced result, we prove in the next Section, that

$$(45) \quad \mathcal{O}_N \subset \{\forall \mathbf{h} \in (\mathbb{L}_\tau^2)^K, (1/2) \|\mathbf{h}\|_\Gamma^2 \leq \|\mathbf{h}\|_N^2 \leq (3/2) \|\mathbf{h}\|_\Gamma^2\}.$$

Now, we prove that, for all  $p_0 \geq 1$ ,  $\mathbb{P}(\mathcal{O}_N^c) \lesssim N^{-p_0}$ .

Indeed, using  $(\mathcal{A}_S)$ , recall that we have set  $\mathfrak{f}_\Gamma := \sup_{t \in [0, \tau]} \|\Gamma(t)^{-1}\|_{\text{op}}$  (see (12)). We have

$$\begin{aligned} \sup_{t \in [0, \tau]} \|\Gamma(t)^{-1/2} \Gamma_N(t) \Gamma(t)^{-1/2} - \text{Id}_K\|_{\text{op}} &\leq \sup_{t \in [0, \tau]} \|\Gamma(t)^{-1}\|_{\text{op}} \sup_{t \in [0, \tau]} \|\Gamma_N(t) - \Gamma(t)\|_{\text{op}} \\ &\leq \mathfrak{f}_\Gamma \sup_{t \in [0, \tau]} \sqrt{\text{Tr}((\Gamma_N(t) - \Gamma(t))^2)}. \end{aligned}$$

Then

$$\begin{aligned} \text{Tr}((\Gamma_N(t) - \Gamma(t))^2) &= \sum_{1 \leq j, k \leq K} \left( \frac{1}{N} \sum_{i=1}^N \{X_{i,j}(t)X_{i,k}(t) - \mathbb{E}[X_{i,j}(t)X_{i,k}(t)]\} \right)^2 \\ &\leq K^2 \max_{1 \leq j, k \leq K} \left( \frac{1}{N} \sum_{i=1}^N \{X_{i,j}(t)X_{i,k}(t) - \mathbb{E}[X_{i,j}(t)X_{i,k}(t)]\} \right)^2. \end{aligned}$$

It follows that

$$\begin{aligned} \mathbb{P}(\mathcal{O}_N^c) &\leq \mathbb{P} \left( \max_{1 \leq j, k \leq K} \sup_{t \in [0, \tau]} \left| \frac{1}{N} \sum_{i=1}^N \{X_{i,j}(t)X_{i,k}(t) - \mathbb{E}[X_{i,j}(t)X_{i,k}(t)]\} \right| > \frac{1}{2K\mathfrak{f}_\Gamma} \right), \\ &\leq \sum_{1 \leq j, k \leq K} \mathbb{P} \left( \sup_{t \in [0, \tau]} \left| \frac{1}{N} \sum_{i=1}^N \{X_{i,j}(t)X_{i,k}(t) - \mathbb{E}[X_{i,j}(t)X_{i,k}(t)]\} \right| > \frac{1}{2K\mathfrak{f}_\Gamma} \right). \end{aligned}$$

The result of Proposition 3 follows immediately from Lemma 6 below.  $\square$

**Lemma 6.** *Let  $p > 2$  and assume that  $(\mathcal{A}_{X,p})$  holds. Let  $j, k \in \{1, \dots, K\}$ , then, there exists a constant  $C_{p,\tau}$  such that, for any constant  $\mathfrak{a}_\tau > 0$ ,*

$$\mathbb{P}_0 := \mathbb{P} \left( \sup_{t \in [0, \tau]} \left| \frac{1}{N} \sum_{i=1}^N (X_{i,j}(t)X_{i,k}(t) - \mathbb{E}[X_{i,j}(t)X_{i,k}(t)]) \right| > \mathfrak{a}_\tau \right) \leq C_{p,\tau} \mathfrak{a}_\tau^{-p} N^{-p/2}.$$

**7.4. Proof of Lemma 6.** We distinguish the proof depending on the case considered in  $(\mathcal{A}_{X,p})$ . We set  $p_0 = p/2$  and, for  $j, k \in \{1, \dots, K\}$  fixed, define

$$(46) \quad Z_N^{j,k}(t) := \frac{1}{N} \sum_{i=1}^N [X_{i,j}(t)X_{i,k}(t) - \mathbb{E}(X_{i,j}(t)X_{i,k}(t))].$$

• **Case (i).** We consider continuous processes  $(X_{i,j})_{1 \leq i \leq N, 1 \leq j \leq K}$  satisfying (13). By Rosenthal Inequality and (13), we obtain that there exists  $a > 1$  and a constant  $c(p, \tau)$  such that

$$\forall N \geq 1, \forall s, t \in [0, \tau], \quad \mathbb{E}[|Z_N^{j,k}(t) - Z_N^{j,k}(s)|^p] \leq c(p, \tau) |t - s|^a \frac{1}{N^{p/2}}.$$

We state the version of The Garsia-Rodemich-Rumsey (1970/71) Lemma [14] given in Jourdain and Pagès (2022), [18].

**Lemma 7.** *Let  $(Y_t^n)_{n \geq 1}$  be a sequence of continuous processes where the processes  $Y^n = (Y_t^n)_{t \in [0, T]}$  are defined on a probability space  $(\Omega, \mathcal{A}, \mathbb{P})$ . Let  $p \geq 1$ . Assume there exists  $a > 1$ , a sequence  $(\delta_n)_{n \geq 1}$  of positive real numbers converging to 0 and a real constant  $C > 0$  such that*

$$\forall n \geq 1, \forall s, t \in [0, T], \quad \mathbb{E}[|Y_t^n - Y_s^n|^p] \leq C |t - s|^a \delta_n^p.$$

*Then there exists a real constant  $C_{p,T} > 0$  such that  $\forall n \geq 1, \mathbb{E} \left[ \sup_{t \in [0, T]} |Y_t^n - Y_0^n|^p \right] \leq C_{p,T} \delta_n^p$ .*

By Lemma 7, there exists a constant  $C_{p,\tau}$  such that

$$\forall N \geq 1, \quad \mathbb{E} \left[ \sup_{t \in [0, \tau]} |Z_N^{j,k}(t) - Z_N^{j,k}(0)|^p \right] \leq C_{p,\tau} \frac{1}{N^{p/2}}.$$

Finally, by the Rosenthal Inequality and  $(\mathcal{A}_{X,p})(a)$ , we get

$$\mathbb{E}[|Z_N^{j,k}(0)|^p] \leq C(p)N^{-p} \{N\mathbb{E}[|X_{1,j}(0)X_{1,k}(0)|^p] + N^{p/2}[\text{Var}(X_{1,j}(0)X_{1,k}(0))]^{p/2}\}.$$

Therefore, there exists another constant  $C_{p,\tau}$ ,

$$(47) \quad \mathbb{E} \left[ \sup_{t \in [0, \tau]} |Z_N^{j,k}(t)|^p \right] \leq C_{p,\tau} \frac{1}{N^{p/2}}.$$

Now by the Markov Inequality,  $\mathbb{P}_0 \leq \mathbf{a}_\tau^{-p} C_{p,\tau} N^{-p/2}$ .  $\square$

• **Case (ii).** We first recall Doob's maximal inequality for martingales: Let  $(M_t)_t$  be a martingale and  $p > 1$ , then it holds that

$$(48) \quad \mathbb{E} \left[ \sup_{t \in [0, \tau]} |M_t|^p \right] \leq \left( \frac{p}{p-1} \right)^p \sup_{t \in [0, \tau]} \mathbb{E}[|M_t|^p].$$

Suppose that  $X_{i,j}(t)$ ,  $1 \leq j \leq K$ , is such that it has moment of any order, independent increments and there exist a continuous deterministic function  $A_j$  such that  $X_{i,j}(t) - A_j(t)$  is a centered martingale. Consider the quantity, for  $1 \leq j, k \leq K$ ,

$$U_1^{j,k}(t) = X_{i,j}(t)X_{i,k}(t) - \mathbb{E}[X_{i,j}(t)X_{i,k}(t)] - A_j(t)(X_{i,k}(t) - A_k(t)) - A_k(t)(X_{i,j}(t) - A_j(t)).$$

Define  $\mathcal{F}_t$  the natural filtration  $\sigma(X_{i,j}(s), s \leq t, 1 \leq j \leq K, 1 \leq i \leq N)$ , we show that  $U_1^{j,k}$  is a martingale. For that we compute for  $t, s \geq 0$

$$\begin{aligned} \mathbb{E}[U_1^{j,k}(t+s)|\mathcal{F}_t] &= \mathbb{E}[X_{1,j}(t+s)X_{1,k}(t+s)|\mathcal{F}_t] - \mathbb{E}[X_{1,j}(t+s)X_{1,k}(t+s)] \\ &\quad - A_j(t+s)(X_{1,k}(t) - A_k(t)) - A_k(t+s)(X_{1,j}(t) - A_j(t)). \end{aligned}$$

We focus on the first term

$$\begin{aligned} \mathbb{E}[X_{1,j}(t+s)X_{1,k}(t+s)|\mathcal{F}_t] &= \mathbb{E}[(X_{1,j}(t+s) - X_{1,j}(t))(X_{1,k}(t+s) - X_{1,k}(t))|\mathcal{F}_t] \\ &\quad + \mathbb{E}[X_{1,j}(t)(X_{1,k}(t+s) - X_{1,k}(t))|\mathcal{F}_t] + \mathbb{E}[X_{1,j}(t+s)X_{1,k}(t)|\mathcal{F}_t] \\ &= \mathbb{E}[(X_{1,j}(t+s) - X_{1,j}(t))(X_{1,k}(t+s) - X_{1,k}(t))] \\ &\quad + X_{1,j}(t)\mathbb{E}[(X_{1,k}(t+s) - X_{1,k}(t))] \\ &\quad + X_{1,k}(t)(\mathbb{E}[X_{1,j}(t+s) - X_{1,j}(t)] + X_{1,j}(t)) \\ &= \mathbb{E}(X_{1,j}(t+s)X_{1,k}(t+s)) + (X_{1,j}(t) - A_j(t))(A_k(t+s) - A_k(t)) \\ &\quad + (X_{1,k}(t) - A_k(t))(A_j(t+s) - A_j(t)) - \mathbb{E}[X_{1,j}(t)X_{1,k}(t)] + X_{1,k}(t)X_{1,j}(t). \end{aligned}$$

where we used the independence of the increments of  $X_{1,j}$ . Plugging this in the latter expression leads to  $\mathbb{E}[U_1^{j,k}(t+s)|\mathcal{F}_t] = U_1^{j,k}(t)$ . Therefore,  $Z_N^{j,k}$  given by (46) can be rewritten

$$\begin{aligned} Z_N^{j,k}(t) &= \frac{1}{N} \sum_{i=1}^N \left\{ U_i^{j,k}(t) + A_j(t)(X_{i,k}(t) - A_k(t)) + A_k(t)(X_{i,j}(t) - A_j(t)) \right\} \\ &= M_{1,N}^{j,k}(t) + A_j(t)M_{2,N}^k(t) + A_k(t)M_{2,N}^j(t) \end{aligned}$$

with  $M_{1,N}^{j,k}(t)$ ,  $M_{2,N}^k(t)$  centered martingales given by

$$M_{1,N}^{j,k}(t) := \frac{1}{N} \sum_{i=1}^N U_i^{j,k}(t), \quad M_{2,N}^k(t) := \frac{1}{N} \sum_{i=1}^N (X_{i,k}(t) - A_k(t))$$

We apply Doob's maximal inequality (48) for  $p > 1$ , and  $C_p = 3^{p-1}(p/p - 1)^p$

$$\mathbb{E} \left[ \sup_{t \in [0, \tau]} |Z_N^{j,k}(t)|^p \right] \leq C_p \left( \sup_{t \in [0, \tau]} \mathbb{E}[|M_{1,N}^{j,k}(t)|^p] + L_\tau \sup_{t \in [0, \tau]} \mathbb{E}[|M_{2,N}^k(t)|^p] + L_\tau \sup_{t \in [0, \tau]} \mathbb{E}[|M_{2,N}^j(t)|^p] \right).$$

The bound of order  $N^{-p/2}$  follows by Rosenthal Inequality, Assumptions (14) and  $(\mathcal{A}_{X,p})$ (a).  $\square$

**7.5. Proof of Inequality (45).** We denote by  $\Gamma(t)^{1/2}$  a symmetric square root of  $\Gamma(t)$ , invertible under  $(\mathcal{A}_S)$ . Let  $\mathbf{h} \in (\mathbb{L}_\tau^2)^K$  such that  $\|\mathbf{h}\|_\Gamma^2 = \int_0^\tau \mathbf{h}(t)^T \Gamma(t) \mathbf{h}(t) dt = 1$ . Then

$$\begin{aligned} \left| \frac{\|\mathbf{h}\|_N^2}{\|\mathbf{h}\|_\Gamma^2} - 1 \right| &= \left| \|\mathbf{h}\|_N^2 - \|\mathbf{h}\|_\Gamma^2 \right| = \left| \int_0^\tau \mathbf{h}(t)^T (\Gamma_N(t) - \Gamma(t)) \mathbf{h}(t) dt \right| \\ &= \left| \int_0^\tau \mathbf{h}(t)^T \Gamma(t)^{1/2} (\Gamma(t)^{-1/2} \Gamma_N(t) \Gamma(t)^{-1/2} - \text{Id}_K) \Gamma(t)^{1/2} \mathbf{h}(t) dt \right| \\ &\leq \sup_{t \in [0, \tau]} \|\Gamma(t)^{-1/2} \Gamma_N(t) \Gamma(t)^{-1/2} - \text{Id}_K\|_{\text{op}} \int_0^\tau |\mathbf{h}(t)^T \Gamma(t) \mathbf{h}(t)| dt \\ &= \sup_{t \in [0, \tau]} \|\Gamma(t)^{-1/2} \Gamma_N(t) \Gamma(t)^{-1/2} - \text{Id}_K\|_{\text{op}}, \end{aligned}$$

using that  $\|\mathbf{h}\|_\Gamma = 1$ . As a consequence  $\sup_{t \in [0, \tau]} \|\Gamma(t)^{-1/2} \Gamma_N(t) \Gamma(t)^{-1/2} - \text{Id}_2\|_{\text{op}} \leq 1/2$  implies for all  $\mathbf{h} \in (\mathbb{L}_\tau^2)^K$ ,  $|\|\mathbf{h}\|_N^2 / \|\mathbf{h}\|_\Gamma^2 - 1| \leq 1/2$ , which gives the result.  $\square$

**7.6. Proof of Lemma 2.** We have

$$\begin{aligned} \mathbb{E} \left[ (\mathbb{W}_m^T \mathbb{W}_m)^2 \right] &= \mathbb{E} \left[ \left( \sum_{j=1}^K \sum_{k=0}^{m_j} \left( \frac{1}{N} \int_0^\tau \varphi_k(t) \sum_{i=1}^N X_{i,j}(t) \sigma(t, \mathbf{X}_i(t)) \varepsilon_i(t) dt \right)^2 \right)^2 \right] \\ &\leq \frac{|\mathbf{m}|}{N^4} \sum_{j=1}^K \sum_{k=0}^{m_j} \mathbb{E} \left[ \left( \sum_{i=1}^N \int_0^\tau \varphi_k(t) X_{i,j}(t) \sigma(t, \mathbf{X}_i(t)) \varepsilon_i(t) dt \right)^4 \right] \\ &\leq \frac{C(4)|\mathbf{m}|}{N^4} \sum_{j=1}^K \sum_{k=0}^{m_j} \left( N \mathbb{E} \left[ \left( \int_0^\tau \varphi_k(t) X_{1,j}(t) \sigma(t, \mathbf{X}_1(t)) \varepsilon_1(t) dt \right)^4 \right] \right. \\ &\quad \left. + N^2 \left\{ \mathbb{E} \left[ \left( \int_0^\tau \varphi_k(t) X_{1,j}(t) \sigma(t, \mathbf{X}_1(t)) \varepsilon_1(t) dt \right)^2 \right]^2 \right\} \right)^2 \\ &\leq \frac{C(4)|\mathbf{m}|}{N^4} \left( \tau^3 c_\varphi K N^{2+\omega} \max_{1 \leq j \leq K} \sup_{t \in [0, \tau]} \mathbb{E} [X_{1,j}^4(t) \sigma^4(t, \mathbf{X}_1(t))] \sup_{t \in [0, \tau]} \mathbb{E} [\varepsilon_1^4(t)] \right. \\ &\quad \left. + |\mathbf{m}| \tau N^2 \left( \max_{1 \leq j \leq K} \sup_{t \in [0, \tau]} \mathbb{E} [X_{1,j}^2(t) \sigma^2(t, \mathbf{X}_1(t))] \sup_{t \in [0, \tau]} \mathbb{E} [\varepsilon_1^2(t)] \right)^2 \right). \end{aligned}$$

where we use Rosenthal inequality with constant  $C(4)$  and Assumption  $(\mathcal{A}_\varphi)$  together with the bounds  $\varphi_k^2(t) \leq L(S_{m_j})$  for  $1 \leq k \leq m_j$ ,  $|\mathbf{m}| \leq N$ , so that

$$\sum_{j=1}^K \sum_{k=0}^{m_j} \int_0^\tau \varphi_k^4(t) dt \leq |\mathbf{m}| \sum_{j=1}^K L(S_{m_j}) \leq c_\varphi |\mathbf{m}| \sum_{j=1}^K m_j^\omega \leq K c_\varphi N^{1+\omega}.$$

Under condition (15) and  $|\mathbf{m}| \leq N$ , we obtain

$$\mathbb{E} \left[ (\mathbb{W}_{\mathbf{m}}^T \mathbb{W}_{\mathbf{m}})^2 \right] \leq C(4) \tau \mathfrak{M}_\sigma^2 (\tau^3 c_\varphi \mathbf{c}_\varepsilon^2 N^{\omega-1} + 1).$$

which gives an order  $O(N^{\omega-1} \vee 1)$  and completes the proof of Lemma 2.  $\square$

**7.7. Proof of inequality (17) of Theorem 1.** Similarly to the previous bound, we write

$$(49) \quad \begin{aligned} \|\widehat{\mathbf{b}}_{\mathbf{m}} - \mathbf{b}\|_\Gamma^2 &= \|\widehat{\mathbf{b}}_{\mathbf{m}} - \mathbf{b}\|_\Gamma^2 \mathbf{1}_{\Lambda_N \cap \mathcal{O}_N} + \|\widehat{\mathbf{b}}_{\mathbf{m}} - \mathbf{b}\|_\Gamma^2 \mathbf{1}_{\Lambda_N \cap \mathcal{O}_N^c} + \|\mathbf{b}\|_\Gamma^2 \mathbf{1}_{\Lambda_N^c} \\ &:= T'_1 + T'_2 + T'_3. \end{aligned}$$

It is straightforward that  $\mathbb{E}[T'_3] = \|\mathbf{b}\|_\Gamma^2 \mathbb{P}(\Lambda_N^c) \lesssim 1/N^p$  for all  $p > 1$ . Now we turn to  $T'_1$ . Let  $\mathbf{b}_{\mathbf{m},\Gamma}$  be the orthogonal projection of  $\mathbf{b}$  on  $\mathbf{S}_{\mathbf{m}}$  w.r.t. the  $\Gamma$ -norm. We have

$$\begin{aligned} \|\widehat{\mathbf{b}}_{\mathbf{m}} - \mathbf{b}\|_\Gamma^2 \mathbf{1}_{\mathcal{O}_N} &= \|\widehat{\mathbf{b}}_{\mathbf{m}} - \mathbf{b}_{\mathbf{m},\Gamma}\|_\Gamma^2 \mathbf{1}_{\mathcal{O}_N} + \|\mathbf{b}_{\mathbf{m},\Gamma} - \mathbf{b}\|_\Gamma^2 \mathbf{1}_{\mathcal{O}_N} \\ &\leq \|\mathbf{b}_{\mathbf{m},\Gamma} - \mathbf{b}\|_\Gamma^2 + 2\|\widehat{\mathbf{b}}_{\mathbf{m}} - \pi_{\mathbf{m}}\mathbf{b}\|_\Gamma^2 \mathbf{1}_{\mathcal{O}_N} + 2\|\mathbf{b}_{\mathbf{m},\Gamma} - \pi_{\mathbf{m}}\mathbf{b}\|_\Gamma^2 \mathbf{1}_{\mathcal{O}_N} \\ &\leq \|\mathbf{b}_{\mathbf{m},\Gamma} - \mathbf{b}\|_\Gamma^2 + 4\|\widehat{\mathbf{b}}_{\mathbf{m}} - \pi_{\mathbf{m}}\mathbf{b}\|_N^2 + 2\|\mathbf{b}_{\mathbf{m},\Gamma} - \pi_{\mathbf{m}}\mathbf{b}\|_\Gamma^2 \mathbf{1}_{\mathcal{O}_N} \end{aligned}$$

Thus, we get

$$\mathbb{E}[T'_1] \leq \|\mathbf{b}_{\mathbf{m},\Gamma} - \mathbf{b}\|_\Gamma^2 + \frac{4}{N} \text{Tr}[\Psi_{\mathbf{m}}^{-1} \Theta_{\mathbf{m}}] + 2\mathbb{E} [\|\mathbf{b}_{\mathbf{m},\Gamma} - \pi_{\mathbf{m}}\mathbf{b}\|_\Gamma^2 \mathbf{1}_{\mathcal{O}_N}].$$

Now, as  $\mathbf{b}_{\mathbf{m},\Gamma}$  and  $\pi_{\mathbf{m}}\mathbf{b}$  belong to  $\mathbf{S}_{\mathbf{m}}$ ,

$$\begin{aligned} \mathbb{E} [\|\mathbf{b}_{\mathbf{m},\Gamma} - \pi_{\mathbf{m}}\mathbf{b}\|_\Gamma^2 \mathbf{1}_{\mathcal{O}_N}] &\leq 2\mathbb{E} [\|\mathbf{b}_{\mathbf{m},\Gamma} - \pi_{\mathbf{m}}\mathbf{b}\|_N^2] = 2\mathbb{E} [\|\pi_{\mathbf{m}}(\mathbf{b} - \mathbf{b}_{\mathbf{m},\Gamma})\|_N^2] \\ &\leq 2\mathbb{E} [\|\mathbf{b} - \mathbf{b}_{\mathbf{m},\Gamma}\|_N^2] = 2\|\mathbf{b} - \mathbf{b}_{\mathbf{m},\Gamma}\|_\Gamma^2. \end{aligned}$$

It follows that

$$\mathbb{E}[T'_1] \leq 5\|\mathbf{b}_{\mathbf{m},\Gamma} - \mathbf{b}\|_\Gamma^2 + \frac{4}{N} \text{Tr}[\Psi_{\mathbf{m}}^{-1} \Theta_{\mathbf{m}}].$$

Let us lastly consider  $T'_2 = \|\widehat{\mathbf{b}}_{\mathbf{m}} - \mathbf{b}\|_\Gamma^2 \mathbf{1}_{\Lambda_N \cap \mathcal{O}_N^c}$  and write

$$T'_2 \leq 2(\|\widehat{\mathbf{b}}_{\mathbf{m}}\|_\Gamma^2 + \|\mathbf{b}\|_\Gamma^2) \mathbf{1}_{\Lambda_N \cap \mathcal{O}_N^c} := T'_{2,1} + T'_{2,2}.$$

Clearly  $\mathbb{E}[T'_{2,2}] \leq \|\mathbf{b}\|_\Gamma^2 \mathbb{P}(\mathcal{O}_N^c) \leq KG_2^2 \|\mathbf{b}\|^2 N^{-p}$ , by using Proposition 2. Analogously, it holds that  $\|\widehat{\mathbf{b}}_{\mathbf{m}}\|_\Gamma^2 \leq KG_2^2 \|\widehat{\mathbf{b}}_{\mathbf{m}}\|^2$ . Now using formula (9), we get  $\|\widehat{\mathbf{b}}_{\mathbf{m}}\|^2 = \|\widehat{\mathbf{B}}_{\mathbf{m}}\|_{2,|\mathbf{m}|}^2 \leq \|\widehat{\Psi}_{\mathbf{m}}^{-1}\|_{\text{op}}^2 \|\widehat{\mathbf{V}}_{\mathbf{m}}\|_{2,|\mathbf{m}|}^2$ . By Proposition 6, on  $\Lambda_N$ , we have

$$\|\widehat{\Psi}_{\mathbf{m}}^{-1}\|_{\text{op}}^2 \leq 4\mathbf{c}_1^2 N^{2c_2}.$$

As a consequence

$$(50) \quad \mathbb{E}[T'_{2,1}] \leq 4\mathbf{c}_1^2 N^{2c_2} \mathbb{E}^{1/2} [\|\mathbf{V}_{\mathbf{m}}\|_{2,|\mathbf{m}|}^4] \mathbb{P}^{1/2}(\mathcal{O}_N^c).$$

By formula (32), we write

$$(51) \quad \|\widehat{\mathbf{V}}_{\mathbf{m}}\|_{2,|\mathbf{m}|}^2 \leq 2 \left( \sum_{j=1}^{|\mathbf{m}|} \langle \mathbf{b}, \Phi_j \rangle_N^2 + \|\mathbb{W}\|_{2,|\mathbf{m}|}^2 \right).$$

By Lemma 2, we have a bound on  $\mathbb{E}^{1/2}[\|\mathbb{W}\|_{2,|\mathbf{m}|}^4] \lesssim N^{(\omega-1)/2} \vee 1$  under  $(\mathcal{A}_{X,p})$ (a). For the first rhs term of (51), we first note that

$$\|\Gamma_N(t)\|_{\text{op}} \leq \frac{1}{N} \sum_{i=1}^N \|\mathbf{X}_i(t)\mathbf{X}_i(t)^T\|_{\text{op}} = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^K X_{i,j}^2(t)$$

and thus if  $(\mathcal{A}_{X,p})$ (a) holds with  $p \geq 8$ ,

$$\mathbb{E}[\|\Gamma_N(t)\|_{\text{op}}^4] \leq K^4 \max_{j \in \{1, \dots, K\}} \sup_{t \in [0, \tau]} \mathbb{E}[X_{1,j}^8(t)] := c_1(\tau, K).$$

Then we get

$$\mathbb{E} \left[ \left( \sum_{j=1}^{|\mathbf{m}|} \langle \mathbf{b}, \Phi_j \rangle_N^2 \right)^2 \right] \leq |\mathbf{m}| \mathbb{E} \left[ \sum_{j=1}^{|\mathbf{m}|} \langle \mathbf{b}, \Phi_j \rangle_N^4 \right]$$

and

$$\begin{aligned} \mathbb{E} [\langle \mathbf{b}, \Phi_j \rangle_N^4] &\leq \mathbb{E} \left[ \int_0^\tau \|\Gamma_N(t)\|_{\text{op}} \|\mathbf{b}(t)\| \|\Phi_j(t)\| dt \right]^4 \\ &\leq \tau^3 \int_0^\tau \mathbb{E}[\|\Gamma_N(t)\|_{\text{op}}^4 \|\mathbf{b}(t)\|^4 \|\Phi_j(t)\|^4] dt \\ &\leq \tau^3 c_1(\tau, K) \left( \sup_{t \in [0, \tau]} \sum_{k=1}^K b_k^2(t) \right)^2 \int_0^\tau \varphi_j^4(t) dt. \end{aligned}$$

Thus,

$$(52) \quad \mathbb{E} \left[ \left( \sum_{j=1}^{|\mathbf{m}|} \langle \mathbf{b}, \Phi_j \rangle_N^2 \right)^2 \right] \leq c_\varphi \tau^3 c_1(\tau, K) \left( \sup_{t \in [0, \tau]} \sum_{k=1}^K b_k^2(t) \right) N^{2+\omega}.$$

Plugging (52) into (50) yields

$$\mathbb{E}[T'_{2,1}] \lesssim N^{2c_2} N^{1+\omega/2} \mathbb{P}^{1/2}(\mathcal{O}_N^c) \lesssim N^{2c_2+1+\omega/2-p/2}.$$

This term is less than  $O(N^{-1})$  for  $p \geq 4c_2 + 4 + \omega$  and  $p \geq 8$ . The bounds on the expectations of  $T'_1, T'_2, T'_3$  yield (17).  $\square$

**7.8. Proof of Proposition 4.** The computation is rather straightforward.

$$\begin{aligned} &\mathbb{E} \left[ \int_0^\tau \left( Y_{N+1}(t) - \tilde{\mathbf{b}}_{\mathbf{m}}^T(t) \mathbf{X}_{N+1}(t) \right)^2 dt \right] \\ &= \mathbb{E} \left[ \int_0^\tau \left( \mathbf{b}^T(t) \mathbf{X}_{N+1}(t) - \tilde{\mathbf{b}}_{\mathbf{m}}^T(t) \mathbf{X}_{N+1}(t) + \sigma(t, \mathbf{X}_{N+1}(t)) \varepsilon_{N+1}(t) \right)^2 dt \right] \\ &= \mathbb{E} \left[ \int_0^\tau \left( (\mathbf{b}(t) - \tilde{\mathbf{b}}_{\mathbf{m}}(t))^T \mathbf{X}_{N+1}(t) \right)^2 dt \right] + \mathbb{E} \left[ \int_0^\tau \sigma^2(t, \mathbf{X}_{N+1}(t)) dt \right] := \text{I} + \text{II} \end{aligned}$$

We compute term I. Denote  $\mathcal{F}_N(\tau)$  the  $\sigma$ -field spanned by  $(Y_1(t), \mathbf{X}_1(t), \dots, Y_N(t), \mathbf{X}_N(t))_{t \in [0, \tau]}$ , it holds

$$\begin{aligned}
\text{I} &= \mathbb{E} \left[ \int_0^\tau (\mathbf{b}(t) - \tilde{\mathbf{b}}_{\mathbf{m}}(t))^T (\mathbf{X}_{N+1}(t) \mathbf{X}_{N+1}(t)^T) (\mathbf{b}(t) - \tilde{\mathbf{b}}_{\mathbf{m}}(t)) dt \right] \\
&= \mathbb{E} \left\{ \mathbb{E} \left[ \int_0^\tau (\mathbf{b}(t) - \tilde{\mathbf{b}}_{\mathbf{m}}(t))^T (\mathbf{X}_{N+1}(t) \mathbf{X}_{N+1}(t)^T) (\mathbf{b}(t) - \tilde{\mathbf{b}}_{\mathbf{m}}(t)) dt \middle| \mathcal{F}_N(\tau) \right] \right\} \\
&= \mathbb{E} \left[ \int_0^\tau (\mathbf{b}(t) - \tilde{\mathbf{b}}_{\mathbf{m}}(t))^T \Gamma(t) (\mathbf{b}(t) - \tilde{\mathbf{b}}_{\mathbf{m}}(t)) dt \right] \\
&= \mathbb{E}[\|\mathbf{b} - \tilde{\mathbf{b}}_{\mathbf{m}}\|_{\Gamma}^2].
\end{aligned}$$

Plugging the value of I in the first equality gives the result of Proposition 4.  $\square$

**7.9. Proof of Proposition 5.** By Proposition 2, we have

$$\begin{aligned}
\inf_{\mathbf{h}=(h_1, \dots, h_K)^T \in S_{\mathbf{m}}} \|\mathbf{h} - \mathbf{b}\|_{\Gamma}^2 &\leq KG_2^2 \inf_{\mathbf{h} \in S_{\mathbf{m}}} \|\mathbf{h} - \mathbf{b}\|^2 = KG_2^2 \inf_{h_j \in S_{m_j}, j=1, \dots, K} \sum_{j=1}^K \|h_j - b_j\|^2 \\
&\leq KG_2^2 \sum_{j=1}^K R_j m_j^{-2s_j}.
\end{aligned}$$

Next, we find, under the assumptions of Theorem 1 with  $\omega = 1$ ,

$$\mathbb{E}[\|\tilde{\mathbf{b}}_{\mathbf{m}} - \mathbf{b}\|_{\Gamma}^2] \leq 5KG_2^2 \sum_{j=1}^K R_j m_j^{-2s_j} + 4\tau \mathfrak{f}_{\Gamma} \mathbf{c}_{\varepsilon} \mathfrak{M}_{\sigma}(1 \wedge \tau c_{\varphi}) \frac{m_1 + \dots + m_K}{N} + \frac{c}{N}.$$

Then choosing  $m_j = CN^{1/(2s_j+1)}$  yields

$$\mathbb{E}[\|\tilde{\mathbf{b}}_{\mathbf{m}} - \mathbf{b}\|_{\Gamma}^2] \lesssim \sum_{j=1}^K N^{-(2s_j)/(2s_j+1)} = O(N^{-2s_{(1)}/(2s_{(1)}+1)}). \quad \square$$

**7.10. Proof of Corollary 2.** Now, let us prove that, if  $\sigma$  is bounded, then  $\text{Tr}[\Psi_{\mathbf{m}}^{-1} \Theta_{\mathbf{m}}] \leq c\tau |\mathbf{m}| \|\sigma\|_{\infty}^2$ . We use the following trick: let  $\mathbf{u} := (u_i)_{1 \leq i \leq |\mathbf{m}|}$  be a vector of i.i.d. centered variables with unit variance, independent of  $(\mathbf{X}_i(t))_{t \in [0, 1], 1 \leq i \leq N}$  and  $(\varepsilon_i(t))_{t \in [0, 1], 1 \leq i \leq N}$ , then for any  $|\mathbf{m}| \times |\mathbf{m}|$  matrix  $C$ , it holds that  $\text{Tr}(C) = \mathbb{E}[\mathbf{u}^T C \mathbf{u}]$ . This implies that

$$\text{Tr}[\Psi_{\mathbf{m}}^{-1} \Theta_{\mathbf{m}}] = \text{Tr}[\Psi_{\mathbf{m}}^{-1/2} \Theta_{\mathbf{m}} \Psi_{\mathbf{m}}^{-1/2}] = \mathbb{E} \left[ \mathbf{u}^T \Psi_{\mathbf{m}}^{-1/2} \Theta_{\mathbf{m}} \Psi_{\mathbf{m}}^{-1/2} \mathbf{u} \right].$$



Setting  $\mathbf{x} = \Psi_{\mathbf{m}}^{-1/2} \mathbf{u}$  and  $\mathbb{E}[\varepsilon(t)\varepsilon(s)] := c(t, s)$  yields

$$\begin{aligned} \mathbf{x}^T \Theta_{\mathbf{m}} \mathbf{x} &= \sum_{1 \leq j, \ell \leq K} \sum_{\substack{1 \leq k \leq m_j \\ 1 \leq p \leq m_\ell}} \mathbf{x}_{m_1 + \dots + m_{j-1} + k} \mathbf{x}_{m_1 + \dots + m_{\ell-1} + p} \\ &\quad \int_0^\tau \int_0^\tau \varphi_k(t) \varphi_p(s) \mathbb{E}[X_{1,j}(t) X_{1,\ell}(s)] \sigma(t, \mathbf{X}_1(t)) \sigma(s, \mathbf{X}_1(s)) c(s, t) ds dt \\ &= \mathbb{E}_{\mathbf{u}} \left[ \left( \int_0^\tau \left( \sum_{j=1}^K h_j(t) X_{1,j}(t) \right) \sigma(t, \mathbf{X}_1(t)) \varepsilon_1(t) \right)^2 dt \right] \\ &\leq \tau \mathbb{E}_{\mathbf{u}} \left[ \int_0^\tau \left( \sum_{j=1}^K h_j(t) X_{1,j}(t) \right)^2 \sigma^2(t, \mathbf{X}_1(t)) \varepsilon_1^2(t) dt \right] \end{aligned}$$

where  $h_j(t) = \sum_{k=1}^{m_j} x_{m_1 + \dots + m_{j-1} + k} \varphi_k(t)$  and  $\mathbb{E}_{\mathbf{u}}[\cdot]$  is the conditional expectation given  $\mathbf{u}$ . Thus we get

$$\mathbf{u}^T \Psi_{\mathbf{m}}^{-1/2} \Theta_{\mathbf{m}} \Psi_{\mathbf{m}}^{-1/2} \mathbf{u} = \mathbf{x}^T \Theta_{\mathbf{m}} \mathbf{x} \leq \tau \|\sigma\|_\infty^2 \int_0^\tau \mathbb{E}_{\mathbf{u}} \left[ \left( \sum_{j=1}^K h_j(t) X_{1,j}(t) \right)^2 \right] c(t, t) dt.$$

Noticing that

$$\int_0^\tau \mathbb{E}_{\mathbf{u}} \left[ \left( \sum_{j=1}^K h_j(t) X_{1,j}(t) \right)^2 \right] dt = \mathbf{x}^T \Psi_{\mathbf{m}} \mathbf{x} = \mathbf{u}^T \Psi_{\mathbf{m}}^{-1/2} \Psi_{\mathbf{m}} \Psi_{\mathbf{m}}^{-1/2} \mathbf{u} = \|\mathbf{u}\|_{2, |\mathbf{m}|}^2,$$

and  $c(t, t) = \mathbb{E}[\varepsilon^2(t)] = 1$ , we obtain, by taking expectation, as  $\mathbb{E}[\|\mathbf{u}\|_{2, |\mathbf{m}|}^2] = |\mathbf{m}|$ ,

$$\text{Tr}(\Psi_{\mathbf{m}}^{-1} \Theta_{\mathbf{m}}) = \text{Tr}[\Psi_{\mathbf{m}}^{-1/2} \Theta_{\mathbf{m}} \Psi_{\mathbf{m}}^{-1/2}] = \mathbb{E} \left[ \mathbf{u}^T \Psi_{\mathbf{m}}^{-1/2} \Theta_{\mathbf{m}} \Psi_{\mathbf{m}}^{-1/2} \mathbf{u} \right] \leq \tau \|\sigma\|_\infty^2 |\mathbf{m}|.$$

Hence, the result.  $\square$

**7.11. Proof of Theorem 3.** *We start by proving (25).* We write the decomposition

$$\|\tilde{\mathbf{b}} - \mathbf{b}\|_N^2 = \|\widehat{\mathbf{b}}_{\widehat{\mathbf{m}}} - \mathbf{b}\|_N^2 \mathbf{1}_{\Lambda_N} + \|\mathbf{b}\|_N^2 \mathbf{1}_{\Lambda_N^c}.$$

The study of the last term is similar to the study of  $T_3$ , see (41)-(42), and yields

$$(53) \quad \mathbb{E}[\|\mathbf{b}\|_N^2 \mathbf{1}_{\Lambda_N^c}] \lesssim \frac{1}{N}$$

thanks to Proposition 6,  $\mathbb{P}(\Lambda_N^c) \lesssim 1/N^p$  for any  $p > 2$ .

For the main term  $\mathbb{E}[\|\widehat{\mathbf{b}}_{\widehat{\mathbf{m}}} - \mathbf{b}\|_N^2 \mathbf{1}_{\Lambda_N}]$ , we recall that  $U_N(\widehat{\mathbf{b}}_{\widehat{\mathbf{m}}}) = -\|\widehat{\mathbf{b}}_{\widehat{\mathbf{m}}}\|_N^2$ . By definition of  $\widehat{\mathbf{b}}_{\widehat{\mathbf{m}}}$ , we have for any  $\mathbf{m} \in \mathcal{M}_N$ , and any  $\mathbf{b}_{\mathbf{m}} \in \mathbf{S}_{\mathbf{m}}$ ,

$$U_N(\widehat{\mathbf{b}}_{\widehat{\mathbf{m}}}) + \text{pen}(\widehat{\mathbf{m}}) \leq U_N(\mathbf{b}_{\mathbf{m}}) + \text{pen}(\mathbf{m}).$$

From (20), we have  $U_N(\mathbf{h}) - U_N(\mathbf{h}^*) = \|\mathbf{h} - \mathbf{b}\|_N^2 - \|\mathbf{h}^* - \mathbf{b}\|_N^2 - 2\nu_N(\mathbf{h} - \mathbf{h}^*)$  with  $\nu_N$  defined by (21), and therefore for any  $\mathbf{m} \in \mathcal{M}_N$ , and any  $\mathbf{b}_{\mathbf{m}} \in \mathbf{S}_{\mathbf{m}}$ , on  $\Lambda_N$

$$\|\widehat{\mathbf{b}}_{\widehat{\mathbf{m}}} - \mathbf{b}\|_N^2 \leq \|\mathbf{b}_{\mathbf{m}} - \mathbf{b}\|_N^2 + \text{pen}(\mathbf{m}) + 2\nu_N(\widehat{\mathbf{b}}_{\widehat{\mathbf{m}}} - \mathbf{b}_{\mathbf{m}}) - \text{pen}(\widehat{\mathbf{m}}).$$

Now we define

$$B_{\mathbf{m}, \mathbf{m}'} = \{\mathbf{h} \in \mathbf{S}_{\mathbf{m}} + \mathbf{S}_{\mathbf{m}'}, \|\mathbf{h}\|_\Gamma = 1\}.$$

We have

$$\mathbb{E}[\|\widehat{\mathbf{b}}_{\widehat{\mathbf{m}}} - \mathbf{b}\|_N^2 \mathbf{1}_{\Lambda_N}] = \mathbb{E}[\|\widehat{\mathbf{b}}_{\widehat{\mathbf{m}}} - \mathbf{b}\|_N^2 \mathbf{1}_{\Lambda_N \cap \mathcal{O}_N}] + \mathbb{E}[\|\widehat{\mathbf{b}}_{\widehat{\mathbf{m}}} - \mathbf{b}\|_N^2 \mathbf{1}_{\Lambda_N \cap \mathcal{O}_N^c}].$$

The term  $\mathbb{E}[\|\widehat{\mathbf{b}}_{\widehat{\mathbf{m}}} - \mathbf{b}\|_N^2 \mathbf{1}_{\Lambda_N \cap \mathcal{O}_N^c}]$  is studied analogously as the previous term  $T_2$ , by noticing that

$$(54) \quad \mathbb{E}[(\mathbb{W}_{\mathbf{m}}^T \mathbb{W}_{\mathbf{m}})^2] \leq \mathbb{E}[(\mathbb{W}_N^T \mathbb{W}_N)^2]$$

and the bound in Lemma 2 applies to the upper bound (as to any  $\mathbf{m}$ ) with constant multiplied by  $K$ . Then using again that  $\mathbb{P}(\mathcal{O}_N^c) \leq c/N^p$  under Assumption  $(\mathcal{A}_{X,p})(a)$ , we get

$$\mathbb{E}[\|\widehat{\mathbf{b}}_{\widehat{\mathbf{m}}} - \mathbf{b}\|_N^2 \mathbf{1}_{\Lambda_N \cap \mathcal{O}_N^c}] \leq \frac{C}{N},$$

for  $p \geq 2\mathbf{c}_2 + 2$ .

Using that, on  $\mathcal{O}_N$ ,  $\forall \mathbf{m}, \mathbf{m}' \in \{1, \dots, N\}^K, \forall \mathbf{h} \in \mathbf{S}_{\mathbf{m}} + \mathbf{S}_{\mathbf{m}'}, \|\mathbf{h}\|_{\Gamma}^2 \leq 2\|\mathbf{h}\|_N^2$ , we obtain, on  $\mathcal{O}_N \cap \Lambda_N$ , the following sequence of inequalities.

$$\begin{aligned} \|\widehat{\mathbf{b}}_{\widehat{\mathbf{m}}} - \mathbf{b}\|_N^2 &\leq \|\mathbf{b}_{\mathbf{m}} - \mathbf{b}\|_N^2 + \text{pen}(\mathbf{m}) + \frac{1}{8}\|\widehat{\mathbf{b}}_{\widehat{\mathbf{m}}} - \mathbf{b}_{\mathbf{m}}\|_{\Gamma}^2 \\ &\quad + 8 \sup_{\mathbf{h} \in B_{\mathbf{m}, \widehat{\mathbf{m}}}} \nu_N^2(\mathbf{h}) - \text{pen}(\widehat{\mathbf{m}}) \\ &\leq \left(1 + \frac{1}{2}\right) \|\mathbf{b}_{\mathbf{m}} - \mathbf{b}\|_N^2 + \text{pen}(\mathbf{m}) + \frac{1}{2}\|\widehat{\mathbf{b}}_{\widehat{\mathbf{m}}} - \mathbf{b}\|_N^2 \\ &\quad + 8 \left( \sup_{\mathbf{h} \in B_{\mathbf{m}, \widehat{\mathbf{m}}}} \nu_N^2(\mathbf{h}) - p(\widehat{\mathbf{m}}, \mathbf{m}) \right) + 8p(\widehat{\mathbf{m}}, \mathbf{m}) - \text{pen}(\widehat{\mathbf{m}}), \end{aligned}$$

where  $p(\widehat{\mathbf{m}}, \mathbf{m}) = \kappa^* \|\sigma\|_{\infty}^2 (|\widehat{\mathbf{m}}| + |\mathbf{m}|)/N$ , where  $\kappa^*$  is a numerical constant (see below). Choosing  $\kappa_0 \geq 8\kappa^*$  implies that  $8p(\widehat{\mathbf{m}}, \mathbf{m}) \leq \text{pen}(\widehat{\mathbf{m}}) + \text{pen}(\mathbf{m})$ . Therefore

$$(55) \quad \mathbb{E}[\|\widehat{\mathbf{b}}_{\widehat{\mathbf{m}}} - \mathbf{b}\|_N^2 \mathbf{1}_{\Lambda_N \cap \mathcal{O}_N}] \leq 3\|\mathbf{b}_{\mathbf{m}} - \mathbf{b}\|_{\Gamma}^2 + 4\text{pen}(\mathbf{m}) + 16\mathbb{E} \left[ \left( \sup_{\mathbf{h} \in B_{\mathbf{m}, \widehat{\mathbf{m}}}} \nu_N^2(\mathbf{h}) - p(\widehat{\mathbf{m}}, \mathbf{m}) \right)_+ \mathbf{1}_{\Lambda_N \cap \mathcal{O}_N} \right].$$

Now we use the following Lemma:

**Lemma 8.** *Under the Assumptions of Theorem 3,*

$$\mathbb{E} \left[ \left( \sup_{\mathbf{h} \in B_{\widehat{\mathbf{m}}, \mathbf{m}}} \nu_N^2(\mathbf{h}) - p(\widehat{\mathbf{m}}, \mathbf{m}) \right)_+ \mathbf{1}_{\widehat{\Lambda}_N \cap \mathcal{O}_N} \right] \leq \frac{C(\tau, \mathbf{f}_{\Gamma}, \|\sigma\|_{\infty})}{N}$$

where  $p(\mathbf{m}, \mathbf{m}') = 3\|\sigma\|_{\infty}^2 \frac{|\mathbf{m}| + |\mathbf{m}'|}{N}$  and  $C(\tau, \mathbf{f}_{\Gamma}, \|\sigma\|_{\infty})$  is a positive constant depending on  $\tau, \mathbf{f}_{\Gamma}$  and  $\|\sigma\|_{\infty}$ .

Gathering Inequalities (53), (54), (55) and the result of Lemma 8 yields Inequality (25) of Theorem 3 with  $\kappa^* = 3$  and  $\kappa \geq 24$ .  $\square$

*Proof of Inequality (26).* First we write, as in the proof of Inequality (17),

$$(56) \quad \begin{aligned} \|\widetilde{\mathbf{b}} - \mathbf{b}\|_{\Gamma}^2 &= \|\widehat{\mathbf{b}}_{\widehat{\mathbf{m}}} - \mathbf{b}\|_{\Gamma}^2 \mathbf{1}_{\Lambda_N \cap \mathcal{O}_N} + \|\widehat{\mathbf{b}}_{\widehat{\mathbf{m}}} - \mathbf{b}\|_{\Gamma}^2 \mathbf{1}_{\Lambda_N \cap \mathcal{O}_N^c} + \|\mathbf{b}\|_{\Gamma}^2 \mathbf{1}_{\Lambda_N^c} \\ &:= T_1'' + T_2'' + T_3''. \end{aligned}$$

The expectation of the last two terms is handled as in the proof of Inequality (17): the term  $T_3''$  is exactly the same as  $T_3'$  and for  $T_2''$ , as the collection of models are such that all  $m'_i$  are less

than  $N$ , the  $\widehat{m}_i$  must be bounded by  $N$  and the  $|\widehat{\mathbf{m}}|$  by  $KN$  before taking expectations. These two terms remain less than  $c/N$ .

For  $T_1''$ , we write that, for any  $\mathbf{m}$  and any  $\mathbf{h} \in S_{\mathbf{m}}$ ,

$$\begin{aligned} T_1'' &\leq 2\|\widehat{\mathbf{b}}_{\widehat{\mathbf{m}}} - \mathbf{h}\|_{\Gamma}^2 \mathbf{1}_{\Lambda_N \cap \mathcal{O}_N} + \|\mathbf{h} - \mathbf{b}\|_{\Gamma}^2 \mathbf{1}_{\Lambda_N \cap \mathcal{O}_N} \\ &\leq 4\|\widehat{\mathbf{b}}_{\widehat{\mathbf{m}}} - \mathbf{h}\|_N^2 \mathbf{1}_{\Lambda_N \cap \mathcal{O}_N} + \|\mathbf{h} - \mathbf{b}\|_{\Gamma}^2 \mathbf{1}_{\Lambda_N \cap \mathcal{O}_N} \\ &\leq 8\|\widehat{\mathbf{b}}_{\widehat{\mathbf{m}}} - \mathbf{b}\|_N^2 \mathbf{1}_{\Lambda_N \cap \mathcal{O}_N} + 8\|\mathbf{b} - \mathbf{h}\|_N^2 \mathbf{1}_{\Lambda_N \cap \mathcal{O}_N} + \|\mathbf{h} - \mathbf{b}\|_{\Gamma}^2 \mathbf{1}_{\Lambda_N \cap \mathcal{O}_N}. \end{aligned}$$

Therefore we get, for any  $\mathbf{m}$  and any  $\mathbf{h} \in S_{\mathbf{m}}$ ,

$$\mathbb{E}[\|\widetilde{\mathbf{b}} - \mathbf{b}\|_{\Gamma}^2] \leq 8\mathbb{E}\left[\|\widehat{\mathbf{b}}_{\widehat{\mathbf{m}}} - \mathbf{b}\|_N^2 \mathbf{1}_{\Lambda_N \cap \mathcal{O}_N}\right] + 10\|\mathbf{b} - \mathbf{h}\|_{\Gamma}^2 + \frac{C'}{N}.$$

Using inequality (25) for the first term and taking the infimum over  $\mathbf{h} \in S_{\mathbf{m}}$  and over  $\mathbf{m}$  yields inequality (26).  $\square$

**7.12. Proof of Lemma 8.** We decompose  $Z_i(\mathbf{h}) := \int_0^{\tau} \mathbf{h}(t)^T \mathbf{X}_i(t) \sigma(t, \mathbf{X}_i(t)) \varepsilon_i(t) dt$  into three parts

$$Z_i(\mathbf{h}) = Z_i^{(1)}(\mathbf{h}) + Z_i^{(2)}(\mathbf{h}) + Z_i^{(3)}(\mathbf{h})$$

where for

$$\Omega_i^X(t) = \left\{ \max_{1 \leq j \leq K} |X_{i,j}(t)| \leq N^{1/4} \right\}, \quad \Omega_i^{\varepsilon}(t) = \left\{ |\varepsilon_i(t)| \leq N^{1/4} \right\}$$

we set

$$\begin{cases} Z_i^{(1)}(\mathbf{h}) = \int_0^{\tau} \mathbf{h}(t)^T \mathbf{X}_i(t) \mathbf{1}_{\Omega_i^X(t)} \sigma(t, \mathbf{X}_i(t)) (\varepsilon_i(t) \mathbf{1}_{\Omega_i^{\varepsilon}(t)} - \mathbb{E}[\varepsilon_i(t) \mathbf{1}_{\Omega_i^{\varepsilon}(t)}]) dt \\ Z_i^{(2)}(\mathbf{h}) = \int_0^{\tau} \mathbf{h}(t)^T \mathbf{X}_i(t) \mathbf{1}_{(\Omega_i^X(t))^c} \sigma(t, \mathbf{X}_i(t)) (\varepsilon_i(t) \mathbf{1}_{\Omega_i^{\varepsilon}(t)} - \mathbb{E}[\varepsilon_i(t) \mathbf{1}_{\Omega_i^{\varepsilon}(t)}]) dt \\ Z_i^{(3)}(\mathbf{h}) = \int_0^{\tau} \mathbf{h}(t)^T \mathbf{X}_i(t) \sigma(t, \mathbf{X}_i(t)) (\varepsilon_i(t) \mathbf{1}_{(\Omega_i^{\varepsilon}(t))^c} - \mathbb{E}[\varepsilon_i(t) \mathbf{1}_{(\Omega_i^{\varepsilon}(t))^c}]) dt \end{cases}$$

As a consequence,

$$\nu_N(\mathbf{h}) = \nu_{N,1}(\mathbf{h}) + \nu_{N,2}(\mathbf{h}) + \nu_{N,3}(\mathbf{h}) \quad \text{with} \quad \nu_{N,k}(\mathbf{h}) = \frac{1}{N} \sum_{i=1}^N Z_i^{(k)}(\mathbf{h}), \quad k = 1, 2, 3.$$

The decomposition of  $Z_i(\mathbf{h})$  gives:

$$\begin{aligned} \mathbb{E} \left[ \left( \sup_{\mathbf{h} \in B_{\widehat{\mathbf{m}}, \mathbf{m}}} \nu_N^2(\mathbf{h}) - p(\widehat{\mathbf{m}}, \mathbf{m}) \right)_+ \mathbf{1}_{\widehat{\Lambda}_N \cap \mathcal{O}_N} \right] &\leq 3\mathbb{E} \left[ \left( \sup_{\mathbf{h} \in B_{\widehat{\mathbf{m}}, \mathbf{m}}} \nu_{N,1}^2(\mathbf{h}) - p(\widehat{\mathbf{m}}, \mathbf{m})/3 \right)_+ \right] \\ &\quad + 3\mathbb{E} \left[ \sup_{\mathbf{h} \in B_{\widehat{\mathbf{m}}, \mathbf{m}}} \nu_{N,2}^2(\mathbf{h}) \right] + 3\mathbb{E} \left[ \sup_{\mathbf{h} \in B_{\widehat{\mathbf{m}}, \mathbf{m}}} \nu_{N,3}^2(\mathbf{h}) \right] \\ &:= \mathbb{T}_1 + \mathbb{T}_2 + \mathbb{T}_3. \end{aligned}$$

We successively bound the three terms, the last two ones being similar.

• **Bound on  $\mathbb{T}_1$ .** We apply the following Talagrand inequality in the form given by Klein and Rio (2005):

**Theorem 4.** *Let  $N \in \mathbb{N}^*$  and  $\mathcal{F}$  a countable class of measurable functions, and  $(Z_i)_{i \in \{1, \dots, N\}}$  a sequence of real random variables. Define, for  $f \in \mathcal{F}$ ,  $\nu_N(f) = \frac{1}{N} \sum_{i=1}^N (f(Z_i) - \mathbb{E}[f(Z_i)])$  a*

linear centered empirical process, and assume that there exist three positive constants  $M$ ,  $H$  and  $v$  such that :

$$\sup_{f \in \mathcal{F}} \|f\|_\infty \leq M, \quad \mathbb{E}[\sup_{f \in \mathcal{F}} |\nu_N(f)|] \leq H \quad \text{and} \quad \sup_{f \in \mathcal{F}} \frac{1}{N} \sum_{i=1}^N \text{Var}(f(Z_i)) \leq v.$$

Then for all  $\alpha > 0$ ,

$$\mathbb{E} \left[ \left( \sup_{f \in \mathcal{F}} |\nu_N(f)|^2 - 2(1 + 2\alpha)H^2 \right)_+ \right] \leq \frac{4}{b} \left( \frac{v}{N} e^{-b\alpha \frac{NH^2}{v}} + \frac{49M^2}{bC(\alpha)^2 N^2} e^{-\frac{\sqrt{2}bC(\alpha)\sqrt{\alpha}}{7} \frac{NH}{M}} \right),$$

where  $b = 1/6$  and  $c(\alpha) = (\sqrt{1 + \alpha} - 1) \wedge 1$ .

First, recall that for  $\mathbf{m} \in \mathcal{M}_N$ , we have  $|\mathbf{m}| \leq KN$ .

$$\mathbb{E} \left[ \left( \sup_{\mathbf{h} \in B_{\widehat{\mathbf{m}}, \mathbf{m}}} \nu_{N,1}^2(\mathbf{h}) - p(\widehat{\mathbf{m}}, \mathbf{m}) \right)_+ \mathbf{1}_{\widehat{\Lambda}_N \cap \mathcal{O}_N} \right] \leq \sum_{\mathbf{m}' \in \mathcal{M}_N} \mathbb{E} \left[ \left( \sup_{\mathbf{h} \in B_{\mathbf{m}, \mathbf{m}'}} \nu_{N,1}^2(\mathbf{h}) - p(\mathbf{m}, \mathbf{m}') \right)_+ \right],$$

we apply the Talagrand inequality to the process  $\nu_{N,1}(\mathbf{h})$  for  $\mathbf{h} \in B_{\mathbf{m}, \mathbf{m}'}$ . Let us denote by  $D(\mathbf{m}, \mathbf{m}')$  the dimension of the space  $\mathbf{S}_{\mathbf{m}} + \mathbf{S}_{\mathbf{m}'}$ , it holds  $D(\mathbf{m}, \mathbf{m}') \leq |\mathbf{m}| + |\mathbf{m}'|$ . Using the isomorphism between  $\mathbf{S}_{\mathbf{m}}$  and  $\mathbf{S}_{|\mathbf{m}|}$  through the basis functions  $\Phi_j$  defined in Section 6.1.1, we consider a family of  $D(\mathbf{m}, \mathbf{m}')$  functions, say  $(\bar{\Phi}_j)_{j=1, \dots, D(\mathbf{m}, \mathbf{m}'})$  which constitutes an orthonormal basis of  $\mathbf{S}_{\mathbf{m}} + \mathbf{S}_{\mathbf{m}'}$  for the scalar product  $\langle \cdot, \cdot \rangle_\tau$ , obtained by Gram-Schmidt orthonormalisation of a  $(\mathbb{L}_\tau^2)^K$  basis deduced from a sub-family of all  $\Phi_j$ 's. Then, by writing  $\mathbf{h} = \sum_{j=1}^{D(\mathbf{m}, \mathbf{m}')} \bar{a}_j \bar{\Phi}_j$ , we have

$$\begin{aligned} & \mathbb{E} \left[ \sup_{\mathbf{h} \in B_{\mathbf{m}, \mathbf{m}'}} |\nu_{N,1}(\mathbf{h})|^2 \right] = \mathbb{E} \left[ \sup_{\mathbf{h} \in B_{\mathbf{m}, \mathbf{m}'}} \left[ \nu_{N,1} \left( \sum_{j=1}^{D(\mathbf{m}, \mathbf{m}')} \bar{a}_j \bar{\Phi}_j \right) \right]^2 \right] \\ & \leq \sum_{j=1}^{D(\mathbf{m}, \mathbf{m}')} \mathbb{E} [\nu_{N,1}^2(\bar{\Phi}_j)] \quad \text{since} \quad \sum_{j=1}^{D(\mathbf{m}, \mathbf{m}')} \bar{a}_j^2 = \|\mathbf{h}\|_\tau^2 = 1 \\ & = \frac{1}{N} \sum_{j=1}^{D(\mathbf{m}, \mathbf{m}')} \text{Var} \left( \int_0^\tau \bar{\Phi}_j^T(t) \mathbf{X}_1(t) \mathbf{1}_{\Omega_1^X(t)} \sigma(t, \mathbf{X}_1(t)) (\varepsilon_1(t) \mathbf{1}_{\Omega_1^\varepsilon(t)} - \mathbb{E}[\varepsilon_1(t) \mathbf{1}_{\Omega_1^\varepsilon(t)}]) dt \right) \\ & \leq \frac{\tau}{N} \sum_{j=1}^{D(\mathbf{m}, \mathbf{m}')} \mathbb{E} \left[ \int_0^\tau (\bar{\Phi}_j^T(t) \mathbf{X}_1(t))^2 \mathbf{1}_{\Omega_1^X(t)} \sigma^2(t, \mathbf{X}_1(t)) (\varepsilon_1(t) \mathbf{1}_{\Omega_1^\varepsilon(t)} - \mathbb{E}[\varepsilon_1(t) \mathbf{1}_{\Omega_1^\varepsilon(t)}])^2 dt \right] \\ & \leq \frac{\tau}{N} \sum_{j=1}^{D(\mathbf{m}, \mathbf{m}')} \|\bar{\Phi}_j\|_\Gamma^2 \|\sigma\|_\infty^2 = \tau \|\sigma\|_\infty^2 \frac{D(\mathbf{m}, \mathbf{m}')}{N} \leq \tau \|\sigma\|_\infty^2 \frac{|\mathbf{m}| + |\mathbf{m}'|}{N} := H^2. \end{aligned}$$

Next, using similar computations we control the term

$$\begin{aligned}
& \sup_{\mathbf{h} \in B_{\mathbf{m}, \mathbf{m}'}} \frac{1}{N} \sum_{i=1}^N \text{Var} \left( \int_0^\tau \mathbf{h}(t)^T \mathbf{X}_i(t) \mathbf{1}_{\Omega_i^X(t)} \sigma(t, \mathbf{X}_i(t)) (\varepsilon_i(t) \mathbf{1}_{\Omega_i^\varepsilon(t)} - \mathbb{E}[\varepsilon_i(t) \mathbf{1}_{\Omega_i^\varepsilon(t)}]) dt \right) \\
&= \sup_{\mathbf{h} \in B_{\mathbf{m}, \mathbf{m}'}} \frac{1}{N} \sum_{i=1}^N \mathbb{E} \left[ \int_0^\tau \mathbf{h}(t)^T \mathbf{X}_i(t) \mathbf{1}_{\Omega_i(X)} \sigma(t, \mathbf{X}_i(t)) (\varepsilon_i(t) \mathbf{1}_{\Omega_i^\varepsilon(t)} - \mathbb{E}[\varepsilon_i(t) \mathbf{1}_{\Omega_i^\varepsilon(t)}]) dt \right]^2 \\
&\leq \tau \sup_{\mathbf{h} \in B_{\mathbf{m}, \mathbf{m}'}} \mathbb{E} \left[ \int_0^\tau (\mathbf{h}(t)^T \mathbf{X}_1(t))^2 \mathbf{1}_{\Omega_1^X(t)} \sigma^2(t, \mathbf{X}_1(t)) (\varepsilon_1(t) \mathbf{1}_{\Omega_1^\varepsilon(t)} - \mathbb{E}[\varepsilon_1(t) \mathbf{1}_{\Omega_1^\varepsilon(t)}])^2 dt \right] \\
&\leq \tau \sup_{\mathbf{h} \in B_{\mathbf{m}, \mathbf{m}'}} \|\mathbf{h}\|_\Gamma^2 \|\sigma\|_\infty^2 = \tau \|\sigma\|_\infty^2 := v.
\end{aligned}$$

Lastly, denoting by  $u(t) := \varepsilon_1(t) \mathbf{1}_{\Omega_1^\varepsilon(t)} - \mathbb{E}[\varepsilon_1(t) \mathbf{1}_{\Omega_1^\varepsilon(t)}]$ ,

$$\begin{aligned}
& \sup_{\mathbf{h} \in B_{\mathbf{m}, \mathbf{m}'}} \sup_{(x, u) \in (\mathbb{L}_\tau^2)^K \times \mathbb{L}_\tau^2} \left| \int_0^\tau \mathbf{h}^T(t) x(t) \mathbf{1}_{\Omega^x(t)} \sigma(t, x(t)) u(t) dt \right|^2 \\
&\leq \tau \sup_{\mathbf{h} \in B_{\mathbf{m}, \mathbf{m}'}} \sup_{(x, u) \in (\mathbb{L}_\tau^2)^K \times \mathbb{L}_\tau^2} \int_0^\tau \left( \sum_{j=1}^K h_j(t) x_j(t) \mathbf{1}_{\Omega^x(t)} \sigma(t, x(t)) u(t) \right)^2 dt \\
&\leq K \tau \sqrt{N} \times \sqrt{N} \|\sigma\|_\infty^2 \sup_{\mathbf{h} \in B_{\mathbf{m}, \mathbf{m}'}} \int_0^\tau \sum_{j=1}^K h_j^2(t) dt \leq \tau K N \|\sigma\|_\infty^2 \mathfrak{f}_\Gamma := M^2
\end{aligned}$$

using Proposition 2 and  $\|\mathbf{h}\|_\Gamma^2 = 1$ .

We apply the Talagrand inequality given in Theorem 4 with  $M^2 = \tau \mathfrak{f}_\Gamma K \|\sigma\|_\infty^2 N$ ,  $v = \|\sigma\|_\infty^2 \tau$  and  $H^2 = \tau \|\sigma\|_\infty^2 (|\mathbf{m}| + |\mathbf{m}'|)/N$  as follows

$$\begin{aligned}
\mathbb{E} \left[ \left( \sup_{\mathbf{h} \in B_{\widehat{\mathbf{m}}, \mathbf{m}}} \nu_{N,1}^2(\mathbf{h}) - p(\widehat{\mathbf{m}}, \mathbf{m}) \right)_+ \mathbf{1}_{\widehat{\Lambda}_N \cap \mathcal{O}_N} \right] &\leq \sum_{\mathbf{m}' \in \mathcal{M}_N} \mathbb{E} \left( \sup_{\mathbf{h} \in B_{\mathbf{m}, \mathbf{m}'}} \nu_{N,1}^2(\mathbf{h}) - p(\mathbf{m}, \mathbf{m}') \right)_+ \\
&\leq \sum_{\mathbf{m}' \in \mathcal{M}_N} \frac{C_1}{N} e^{-C_2 \sqrt{|\mathbf{m}| + |\mathbf{m}'|}} \leq \frac{C}{N},
\end{aligned}$$

for some positive constant  $C$  and with  $p(\mathbf{m}, \mathbf{m}') = 3 \|\sigma\|_\infty^2 \frac{|\mathbf{m}| + |\mathbf{m}'|}{N}$ , for  $\alpha = 1/4$ .

• Now, we turn to the bound on  $\mathbb{T}_2$ . We prove that  $\mathbb{E} \left( \sup_{\mathbf{h} \in B_{\widehat{\mathbf{m}}, \mathbf{m}}} \nu_{N,2}^2(\mathbf{h}) \right) \leq C/N$ . Recall that for all  $\mathbf{m}$ ,  $\mathbf{S}_\mathbf{m} \subset \mathbf{S}_N$ .

$$\begin{aligned}
\mathbb{E} \left[ \sup_{\mathbf{h} \in B_{\widehat{\mathbf{m}}, \mathbf{m}}} |\nu_{N,2}(\mathbf{h})|^2 \right] &\leq \mathbb{E} \left[ \sup_{\mathbf{h} \in \mathcal{S}_N, \|\mathbf{h}\|_{\Gamma}=1} |\nu_{N,2}(\mathbf{h})|^2 \right] = \mathbb{E} \left[ \sup_{\mathbf{h} \in \mathcal{S}_N, \|\mathbf{h}\|_{\Gamma}=1} \left[ \nu_{N,2} \left( \sum_{j=1}^{KN} a_j \Phi_j \right) \right]^2 \right] \\
&= \sup_{\mathbf{h} \in \mathcal{S}_N, \|\mathbf{h}\|_{\Gamma}=1} \|\mathbf{h}\|^2 \mathbb{E} \left[ \sum_{j=1}^{KN} \nu_{N,2}^2(\Phi_j) \right] \\
&\leq \frac{\mathfrak{f}_{\Gamma}}{N} \sum_{j=1}^{KN} \mathbb{E} \left[ \int_0^{\tau} (\Phi_j^T(t) \mathbf{X}_1(t))^2 \mathbf{1}_{\Omega_1^c(X)} \sigma^2(t, \mathbf{X}_1(t)) (\varepsilon_1(t) \mathbf{1}_{\Omega_1^c(t)} - \mathbb{E}[\varepsilon_1(t) \mathbf{1}_{\Omega_1^c(t)}])^2 dt \right] \\
&\leq \frac{\mathfrak{f}_{\Gamma}}{N} \tau \|\sigma\|_{\infty}^2 \sum_{j=1}^{|\mathbf{m}|+|\widehat{\mathbf{m}}|} \mathbb{E} \left[ \int_0^{\tau} (\Phi_j^T(t) \mathbf{X}_1(t))^2 \mathbf{1}_{(\Omega_1^X(t))^c} dt \right],
\end{aligned}$$

using that  $\mathbb{E}[\varepsilon_1^2(t) \mathbf{1}_{\Omega_1^c(t)}] \leq 1$ . Then as we can write  $\sum_{j=1}^{KN} (\Phi_j^T(t) \mathbf{X}_1(t))^2 = \sum_{j=1}^K \sum_{k=1}^N (\varphi_k(t) X_{1j}(t))^2$ , we get

$$\begin{aligned}
\mathbb{E} \left[ \sup_{\mathbf{h} \in B_{\widehat{\mathbf{m}}, \mathbf{m}}} |\nu_{N,2}(\mathbf{h})|^2 \right] &\leq \frac{\mathfrak{f}_{\Gamma}}{N} \tau \|\sigma\|_{\infty}^2 \sum_{j=1}^K \sum_{k=1}^N \mathbb{E} \left[ \int_0^{\tau} (\varphi_k(t) X_{1j}(t))^2 \mathbf{1}_{(\Omega_1^X(t))^c} dt \right] \\
&\leq \frac{\mathfrak{f}_{\Gamma}}{N} \tau \|\sigma\|_{\infty}^2 \sum_{j=1}^K \mathbb{E} \left[ \int_0^{\tau} \left( \sup_{t \in [0, \tau]} \sum_{k=1}^N \varphi_k^2(t) \right) X_{1j}^2(t) \mathbf{1}_{(\Omega_1^X(t))^c} dt \right] \\
&\leq \frac{\mathfrak{f}_{\Gamma}}{N} \tau \|\sigma\|_{\infty}^2 L(S_N) \mathbb{E} \left[ \int_0^{\tau} \sum_{j=1}^K X_{1j}^2(t) \mathbf{1}_{\sum_{j=1}^K X_{1,j}^2(t) > N^{1/2}} dt \right] \\
&\leq \frac{\mathfrak{f}_{\Gamma}}{N} \tau^2 \|\sigma\|_{\infty}^2 c_{\varphi}(KN) \frac{\sup_{t \in [0, \tau]} \mathbb{E}[(\sum_{j=1}^K X_{1j}^2(t))^{1+q}]}{N^{q/2}} \\
&\leq \mathfrak{f}_{\Gamma} \tau^2 \|\sigma\|_{\infty}^2 c_{\varphi} K^{2+q} \frac{\max_{1 \leq j \leq K} \sup_{t \in [0, \tau]} \mathbb{E}[X_{1j}^{2+2q}(t)]}{N^{q/2}} \leq \frac{C}{N}
\end{aligned}$$

as soon as  $q \geq 2$  and Assumption  $(\mathcal{A}_{X,p})(a)$  holds for  $p \geq 6$ .

- The bound on  $\mathbb{T}_3$  is obtained in a similar way as the second one and is omitted. It requires  $\sup_{t \in [0, \tau]} \mathbb{E}[\varepsilon_1^6(t)] < +\infty$ .

Gathering the three bounds gives the result of Lemma 8.  $\square$